



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12Q	A2	(11) International Publication Number: WO 99/06587 (43) International Publication Date: 11 February 1999 (11.02.99)
(21) International Application Number: PCT/EP98/04836 (22) International Filing Date: 3 August 1998 (03.08.98) (30) Priority Data: 97113319.4 1 August 1997 (01.08.97) EP (71) Applicant (for all designated States except US): MORPHOSYS GESELLSCHAFT FÜR PROTEINOPTIMIERUNG AG [DE/DE]; Am Klopferspitz 19, D-82152 Martinsried (DE). (72) Inventors; and (75) Inventors/Applicants (for US only): RUDERT, Fritz [DE/DE]; Josef-Retzer-Strasse 36, D-81241 München (DE). GE, Liming [CN/DE]; Portiastrasse 12, D-81545 München (DE). ILAG, Vic [PH/DE]; Knorrstrasse 85, D-89897 München (DE). (74) Agent: VOSSIUS & PARTNER; Siebertstrasse 4, D-81675 München (DE).		(81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: NOVEL METHOD AND PHAGE FOR THE IDENTIFICATION OF NUCLEIC ACID SEQUENCES ENCODING MEMBERS OF A MULTIMERIC (POLY)PEPTIDE COMPLEX <div style="display: flex; justify-content: space-around;"> <div data-bbox="365 1150 755 1753"> <p>General description of the polyphage principle</p> </div> <div data-bbox="836 1150 1226 1753"> <p>General description of the polyphage principle (cont.)</p> </div> </div> (57) Abstract <p>The present invention relates to methods for the identification of nucleic acid sequences encoding members of a multimeric (poly)peptide complex by screening for polyphage particles. Furthermore, the invention relates to products and uses thereof for the identification of nucleic acid sequences in accordance with the present invention.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

NOVEL METHOD AND PHAGE FOR THE IDENTIFICATION OF NUCLEIC ACID SEQUENCES ENCODING MEMBERS OF A MULTIMERIC (POLY)PEPTIDE COMPLEX

The present invention relates to methods for the identification of nucleic acid sequences encoding members of a multimeric (poly)peptide complex by screening for polyphage particles. Furthermore, the invention relates to products and uses thereof for the identification of nucleic acid sequences in accordance with the present invention.

Since its first conception by Ladner in 1988 (WO88/06630), the principle of displaying repertoires of proteins on the surface of phage has experienced a dramatic progress and has resulted in substantial achievements. Initially proposed as display of single-chain Fv (scFv) fragments, the method has been expanded to the display of bovine pancreatic trypsin inhibitor (BPTI) (WO90/02809), human growth hormone (WO92/09690), and of various other proteins including the display of multimeric proteins such as Fab fragments (WO91/17271; WO92/01047).

A Fab fragment consists of a light chain comprising a variable and a constant domain (VL-CL) non-covalently binding to a heavy chain comprising a variable and constant domain (VH-CH1). In Fab display one of the chains is fused to a phage coat protein, and thereby displayed on the phage surface, and the second is expressed in free form, and on contact of both chains, the Fab assembles on the phage surface.

Various formats have been developed to construct and screen Fab phage-display libraries. In its simplest form, just one repertoire, e. g. of heavy chains, is encoded on the phage or phagemid vector. A corresponding light chain, or a repertoire of light chains, is expressed separately. The Fab fragments assemble either inside a host cell, if the light chain is co-expressed from a plasmid, or outside the cell in the medium, if a collection of secreted phage particles each displaying a heavy chain is contacted with the light chain(s) expressed from a different host cell. By screening such Fab libraries, just the information about the heavy chain encoded on the phage or phagemid vector is retrievable, since that vector is packaged in the phage particle. By reverting the format and displaying a library of light chains, and

assembling Fab fragments by co-expressing or adding one or more of the heavy chains identified in the first round, corresponding light chain-heavy chain pairs can be identified.

To avoid that multi-step procedure, both repertoires may be cloned into one phage or phagemid vector, one chain expressible as a fusion with at least part of a phage coat protein, the second expressible in free form. After selection, the phage particle will contain the sequence information about both chains of the selected Fab fragments. The disadvantage of such a format is that the overall complexity of the library is limited by transformation efficiency. Therefore, the library size will usually not exceed 10^{10} members.

For various applications, a library size of up to 10^{14} would be advantageous. Therefore, methods of using site-specific recombination, either based on the Cre/lox system (WO92/20791) or on the att λ system (WO 95/21914) have been proposed. Therein, two collection of vectors are sequentially introduced into host cells. By providing the appropriate recombination sites on the individual vectors, recombination between the vectors can be achieved by action of an appropriate recombinase or integrase, achieving a combinatorial library, the overall library size being the product of the sizes of the two individual collections. The disadvantages of the Cre/lox system are that the recombination event is not very efficient, it leads to different products and is reversible. The att λ system leads to a defined product, however, it creates one very large plasmid which has a negative impact on the production of phages. Furthermore, the action of recombinase or integrase most likely leads to undesired recombination events.

Thus, the technical problem underlying the present invention is to develop a simple, reliable system which enables the simultaneous identification of members of a multimeric (poly)peptide complex, such as the identification of heavy and light chain of a Fab fragment, in phage display systems.

The solution to this technical problem is achieved by providing the embodiments characterized in the claims. Accordingly, the present invention allows to easily create and screen large libraries of multimeric (poly)peptide complexes for properties such as binding to a target, as in the case of screening Fab fragment libraries, or such as enzymatic activity, as in the case of libraries of multimeric enzymes. The technical approach of the present invention, i.e. the retrieval of information about two members of a multimeric (poly)peptide complex

encoded on two different vectors without requiring a recombination event, is neither provided nor suggested by the prior art.

Accordingly, the present invention relates to a method for identifying a combination of nucleic acid sequences encoding two members of a multimeric (poly)peptide complex with a predetermined property, said combination being contained in a combinatorial library of phage particles displaying a multitude of multimeric (poly)peptides complexes, said method being characterized by screening or selecting for polyphage particles that contain said combination.

Surprisingly, it has been achieved by the present invention that the phenomenon of polyphages can be used to co-package the genetic information of two or more members of multimeric (poly)peptide complexes in a phage display system. The occurrence of polyphage particles has been observed 30 years ago (Salivar et al., Virology 32 (1967) 41-51), where it was described that approximately 5% of a phage population form particles which are longer than unit length and which contain two or more copies of phage genomic DNA. They occur naturally when a newly forming phage coat encapsulates two or more single-stranded DNA molecules. In specific cases, it has been seen that co-packaging of phage and phagemids or single-stranded plasmid vectors takes place as well (Russel and Model, J. Virol. 63 (1989) 3284-3295). Despite of occasional scientific articles about the morphogenesis of polyphage particles, a practical application has never been discussed or even been mentioned. In WO92/20791 in example 26, a model experiment for a combinatorial Fab display library expressed from separate vectors is presented. However, there is only a screening process for either of the two vectors described. Thus, the prior art teaches away from screening for the simultaneous presence of two vectors in a polyphage particle.

In the context of the present invention, the term "multimeric (poly)peptide complex" refers to a situation where two or more (poly)peptide(s) or protein(s), the "members" of said multimeric complex, can interact to form a complex. The interaction between the individual members will usually be non-covalent, but may be covalent, when post-translational modification such as the formation of disulphide-bonds between any two members occurs. Examples for "multimeric (poly)peptide complexes" comprise structures such as fragments derived from immunoglobulins (e. g. Fv, disulphide-linked Fv (dsFv), Fab fragments), fragments derived from other members of the immunoglobulin superfamily (e.g. α , β -

heterodimer of the T-cell receptor), and fragments derived from homo-or heterodimeric receptors or enzymes. In phage display, one of said members is fused to at least part of a phage coat protein, whereby that member is displayed on, and assembly of the multimeric complex takes place at, the phage surface. A "combinatorial phage library" is produced by randomizing at least two members of said multimeric (poly)peptide complex at least partially on the genetic level to create two libraries of genetically diverse nucleic acid sequences in appropriate vectors, by combining the libraries in appropriate host cells and by achieving co-expression of said at least two libraries in a way that a library of phage particles is produced wherein each particle displays one of the possible combinations out of the two libraries.

By screening such a combinatorial phage library displaying multimeric (poly)peptide complexes for a predetermined property, a collection of phage particles will be identified. Partially, these particles will just contain the genetic information of one of the members of the multimeric complex. The inventive principle of the present invention is the screening step for polyphage particles containing the genetic information of a combination of library members.

Furthermore, the present invention relates to a method for identifying a combination of nucleic acid sequences encoding two members of a multimeric (poly)peptide complex with a predetermined property, said combination being contained in a combinatorial library of phage particles displaying a multitude of multimeric (poly)peptides complexes, comprising the steps of

- (a) providing a first library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences, each encoding a fusion protein of a first member of a multimeric (poly)peptide complex fused to at least part of a phage coat protein, said fusion protein thereby being able to be directed to, and displayed at, the phage surface, wherein said vector molecules are able to be packaged in a phage particle and carry or encode a first selectable and/or screenable property;
- (b) providing a second library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences, each encoding a second member of a multimeric (poly)peptide complex, wherein the vector molecules of said second library are able to be packaged in a phage particle and carry

or encode a second selectable and/or screenable property different from said first property;

- (c) optionally, providing nucleic acid sequences encoding further members of a multimeric (poly)peptide complex;
- (d) expressing members of said libraries of recombinant vectors mentioned in steps (a), (b), and optionally nucleic acid sequences mentioned in step (c), in appropriate host cells under appropriate conditions, so that a combinatorial library of phage particles each displaying a multimeric (poly)peptide complex is produced;
- (e) identifying in said library of phage particles a collection of phages displaying multimeric (poly)peptide complexes having said predetermined property;
- (f) identifying in said collection polyphage particles simultaneously containing recombinant vector molecules encoding a first and a second member of said multimeric (poly)peptide complex by screening or selecting for the simultaneous presence or generation of said first and second selectable and/or screenable property;
- (g) optionally, carrying out further screening and/or selection steps or repeating steps (a) to (f);
- (h) identifying said combination of nucleic acid sequences.

Optionally, further members of said multimeric complex may be provided in the case of ternary, quaternary or higher (poly)peptide complexes. These further members may, for example, be co-expressed from one of the phage or phagemid vectors or from a separate vector such as a plasmid. Even libraries of such further members could be employed in which case further screenable or selectable properties would have to be introduced on the corresponding vectors. Alternatively, such further libraries could be contained in said first or second libraries of recombinant vector molecules. In another option, further screening and/or selection steps or a repetition of the individual steps can be carried out, to optimize the result of obtaining and identifying said nucleic acid sequences.

Furthermore, the present invention relates to a method for identifying a combination of nucleic acid sequences encoding two members of a multimeric (poly)peptide complex with a predetermined property, said combination being contained in a combinatorial library of phage particles displaying a multitude of multimeric (poly)peptides complexes, comprising the steps of

- (a) expressing in appropriate host cells under appropriate conditions

- (aa) genetically diverse nucleic acid sequences contained in a first library of recombinant vector molecules, said nucleic acid sequences comprising a variety of nucleic acid sequences, each encoding a fusion protein of a first member of a multimeric (poly)peptide complex fused to at least part of a phage coat protein, said fusion protein thereby being able to be directed to and displayed at the phage surface, wherein said vector molecules are able to be packaged in a phage particle and carry or encode a first selectable and/or screenable property;
- (ab) genetically diverse nucleic acid sequences contained in a second library of recombinant vector molecules, said nucleic acid sequences comprising a variety of nucleic acid sequences, each encoding a second member of a multimeric (poly)peptide complex, wherein the vector molecules are able to be packaged in a phage particle and carry or encode a second selectable and/or screenable property different from said first property;
- (ac) optionally, nucleic acid sequences encoding further members of a multimeric (poly)peptide complex,
so that a combinatorial library of phage particles each displaying a multimeric (poly)peptide complex is produced;
- (b) identifying in said library of phage particles a collection of phages displaying multimeric (poly)peptide complexes having said predetermined property;
- (c) identifying in said collection polyphage particles simultaneously containing recombinant vector molecules encoding a first and a second member of said multimeric (poly)peptide complex by screening or selecting for the simultaneous presence or generation of said first and second selectable and/or screenable property;
- (d) optionally, carrying out further screening and/or selection steps or repeating steps (a) to (c);
- (e) identifying said combination of nucleic acid sequences.

In a preferred embodiment of the method of the present invention, the vectors of said first and said second library are a combination of a phage vector and a phagemid vector.

In a further preferred embodiment of the method of the present invention, the vectors of said first and said second library are a combination of two phagemid vectors, said appropriate conditions comprising complementation of phage genes by a helper phage.

In a most preferred embodiment of the method of the present invention said two phagemid vectors are compatible.

The term "compatibility" refers to a property of two phagemids to be able to coexist in a host cell. Incompatibility is connected to the presence of incompatible plasmid origins of replication belonging to the same incompatibility group. An example for compatible plasmid origins of replication is the high-copy number origin ColE1 and the low-copy number origin p15A.

Therefore, in a further preferred embodiment of the method of the present invention, said two phagemid vectors comprise a ColE1 and a p15A plasmid origin of replication.

In a most preferred embodiment of the method of the present invention, said two phagemid vectors comprise a ColE1 and a mutated ColE1 origin.

It could be shown, that two phagemids both having a ColE1-derived plasmid origin of replication can coexist in a cell as long as one of the ColE1 origins carries a mutation.

Particularly preferred is a method, wherein said vectors and/or said helper phage comprise different phage origins of replication.

Most preferred is an embodiment of the method of the present invention, wherein said phage vector, said phagemid vector(s) and/or said helper phage are interference resistant.

The term "interference" refers to a property that phagemids inhibit the production of progeny phage particles by interfering with the replication of the DNA of the phage. "Interference resistance" is a property which overcomes this problem. It has been found that mutations in the intergenic region and/or in gene II contribute to interference resistance (Enea and Zinder, Virology 122 (1982), 222-226; Russel et al., Gene 45 (1986) 333-338). It was identified that phages called IR1 and IR2 (Enea and Zinder, Virology 122 (1982), 222-226), and mutants derived therefrom such as R176 (Russel and Model, J. Bacteriol. 154 (1983) 1064-1076), R382, R407 and R408 (Russel et al., Gene 45 (1986) 333-338) and R383 (Russel and Model, J. Virol. 63 (1989) 3284-3295) are interference resistant by carrying mutations in the untranslated region upstream of gene II and in the gene II coding region.

Therefore, in a preferred embodiment of the method of the present invention, said phage vector, said phagemid vector(s) and/or said helper phage have mutations in the phage intergenic region(s), preferably in positions corresponding to position 5986 of f1, and/or in gene II, preferably in positions corresponding to position 143 of f1.

In a most preferred embodiment said phage vector, said phagemid vector(s) and/or said helper phage are, or are derived from, IR1 mutants such as R176, R382, R383, R407, R408, or from IR2 mutants.

In a further embodiment of the method of the invention, said vectors and/or said helper phage comprise hybrid nucleic acid sequences of f1, fd, and/or M13 derived sequences.

In the context of the present invention, the term "hybrid nucleic sequences" refers to vector elements which comprise sequences originating from different phage(mid) vectors.

Surprisingly, it has been found that a vector constructed combining a part derived from fd phage and a second part derived from R408, a derivative of f1 phages, is interference resistant and additionally, gives predominantly polyphage particles.

Therefore, a most preferred embodiment of the method of the present invention relates to a vector which is, or is derived from, fpep3_1B-IR3seq with the sequence listed in Figure 4.

In a yet further preferred embodiment of the method according to the present invention, said derivative is a phage comprising essentially the phage origin or replication from fpep3_1B-IR3seq, the gene II from fpep3_1B-IR3seq, or a combination of said phage origin of replication and said gene II.

The invention relates in an additional preferred embodiment to a method, wherein said derivative is a phagemid comprising essentially the phage origin or replication from fpep3_1B-IR3seq, the gene II from fpep3_1B-IR3seq, or a combination of said phage origin of replication and said gene II.

The invention relates in a further preferred embodiment to a method, wherein said derivative is a helper phage comprising essentially the phage origin or replication from fpep3_1B-

IR3seq, the gene II from fpep3_1B-IR3seq, or a combination of said phage origin of replication and said gene II.

Most preferred is an embodiment of the method of the invention, wherein said derivatives comprise the combined fd/f1 origin including the mutation G5737>A (2976 in fpep3_1B-IR3seq), and/or the mutations G343>A (3989) in gII, and G601>T (4247) in gII/X.

The formation of polyphage particles has been examined in more detail by different groups. It was found that amber mutations in genes VII and IX lead to the amplified production of infectious polyphage particles (Lopez and Webster, Virology 127 (1983) 177-193). A couple of mutants in gene VII (R68, R100) and in gene IX (N18) were identified and further characterized.

Accordingly, in a preferred embodiment of the method of the present invention, the gene VII contained in any of said vectors contains an amber mutation, and most preferably, said mutation is identical to those found in phage vectors R68 or R100.

Further preferred is an embodiment, wherein the gene IX contained in any of said vectors contains an amber mutation, and most preferably said mutation is identical to that found in phage vector N18.

Several phage coat proteins have been used in displaying foreign proteins including the gene III protein (gIIIp), gVIp, and gVIIIp.

In a preferred embodiment of the method of the present invention, said phage coat protein is gIIIp or gVIIIp.

In a particularly preferred embodiment of the method of the present invention, said phage particles are infectious by having a full-length copy of gIIIp.

The gIIIp is a protein comprising three domains. The C-terminal domain is responsible for membrane insertion, the two N-terminal domains are responsible for binding to the F pilus of *E. coli* (N2) and for the infection process (N1).

In a most preferred embodiment of the method of the invention, said phage particles are non-infectious by having no full-length copy of gIIIp, said fusion protein being formed with a truncated version of gIIIp, wherein the infectivity can be restored by interaction of the

displayed multimeric (poly)peptide complexes with a corresponding partner coupled to an infectivity-mediating particle.

In the context of the present invention, the term "infectivity-mediating particle" (IMP) refers to a construct comprising either the N1 domain or the N1-N2 domain. On interaction with a non-infectious phage lacking said domains, infectivity of the phage particles can be restored. The interaction between the non-infectious phage and the IMP can be mediated by a ligand fused to the IMP, which can bind to a partner displayed on the phage. By screening a non-infectious phage display library against a target ligand-IMP construct, restoration of infectivity can be used to select target-binding library members.

In a further preferred embodiment of the method of the invention, said truncated gIIIp comprises the C-terminal domain of gIIIp.

In a yet preferred embodiment of the method of the invention, said truncated gIIIp is derived from phage fCA55.

In addition to the work by Lopey and Webster cited above, Crissman and Smith (Virology 132 (1984) 445-455) could show, that the phage fCA55 which has a large deletion in gene III removing the N-terminal domains and a large part of the C-terminal domain leads exclusively to the formation of polyphages.

Particularly preferred is an embodiment of the method of the invention, wherein said predetermined property is binding to a target.

In a preferred embodiment of the method of the invention, said multimeric (poly)peptide complex is a fragment of an immunoglobulin superfamily member.

In a most preferred embodiment of the method of the invention, said multimeric (poly)peptide complex is a fragment of an immunoglobulin.

In a further most preferred embodiment of the method of the invention, said fragment is an Fv, dsFv or Fab fragment.

An additional preferred embodiment of the present invention relates to a method, wherein said predetermined property is the activity to perform or to catalyze a reaction.

In a preferred embodiment of the method of the invention, said multimeric (poly)peptide complex is an enzyme.

In a most preferred embodiment of the method of the invention, said multimeric (poly)peptide complex is a fragment of a catalytic antibody.

In a further most preferred embodiment of the method of the invention, said fragment is an Fv, dsFv or Fab fragment.

An additional preferred embodiment of the invention relates to a method, wherein selectable and/or screenable property is the transactivation of transcription of a reporter gene such as beta-galactosidase, alkaline phosphatase or nutritional markers such as his3 and leu, or resistance genes giving resistance to an antibiotic such as ampicillin, chloramphenicol, kanamycin, zeocin, neomycin, tetracycline or streptomycin.

In a most preferred embodiment of the method of the invention, said generation of said first and second screenable and/or selectable property is achieved after infection of appropriate host cells by said collection of phage particles.

Particularly preferred is a method, wherein said identification of said nucleic acid sequences is effected by sequencing.

Further preferred is a method, wherein said host cells are E.coli XL-1 Blue, K91 or derivatives, TG1, XL1kann or TOP10F.

An additional preferred embodiment of the invention relates to a polyphage particle which

(a) contains

(i) a first recombinant vector molecule that comprises a nucleic acid sequence, which encodes a fusion protein of a first member of a multimeric (poly)peptide complex

fused to at least part of a phage coat protein, and that carries or encodes a first selectable and/or screenable property, and

(ii) a second recombinant vector molecule that comprises a nucleic acid sequence, which encodes a second member of a multimeric (poly)peptide complex, and that carries or encodes a second selectable and/or screenable property different from said first property;

and (b) displays said multimeric (poly)peptide complex at its surface.

A most preferred embodiment of the invention relates to a polyphage particle, wherein said phage coat protein is the gIIIp.

A further preferred embodiment of the present invention relates to a polyphage particle which is infectious by having a full-length copy of gIIIp present, either in said fusion protein, or in an additional wild-type copy.

Additionally, the invention relates to a polyphage particle which is non-infectious by having no full-length copy of gIIIp, said fusion protein being formed with a truncated version of gIIIp, wherein the infectivity can be restored by interaction of the displayed multimeric (poly)peptide complex with a corresponding partner coupled to an infectivity-mediating particle.

Most preferably, the invention relates to the phage vector fpep3_1B-IR3seq with the sequence listed in Figure 4.

Additionally preferred, the invention relates to a phage vector derived from phage vector fpep3_1B-IR3seq comprising essentially the phage origin or replication from fpep3_1B-IR3seq, the gene II from fpep3_1B-IR3seq, or a combination of said phage origin of replication and said gene II.

Further preferred is an embodiment of the invention, which relates to a phagemid vector derived from phage vector fpep3_1B-IR3seq comprising essentially the phage origin or replication from fpep3_1B-IR3seq, the gene II from fpep3_1B-IR3seq, or a combination of said phage origin of replication and said gene II.

Preferably, the invention relates to a helper phage vector derived from phage vector fpep3_1B-IR3seq comprising essentially the phage origin or replication from fpep3_1B-IR3seq, the gene II from fpep3_1B-IR3seq, or a combination of said phage origin of replication and said gene II.

Additionally preferred is an embodiment, said derivatives comprise the combined fd/fl origin including the mutation G5737>A (2976 in fpep3_1B-IR3seq), and/or the mutations G343>A (3989) in gII, and G601>T (4247) in gII/X.

Further preferred is the use of any of the vectors according to the present invention in the generation of polyphage particles containing a combination of at least two different vectors.

Most preferred is the use of vectors of the invention, wherein said combination of different vectors comprises nucleic acid sequences encoding members of a multimeric (poly)peptide complex.

Further preferred in the present invention is the use of vectors, wherein said combination of different vectors comprises nucleic acid sequences encoding interacting (poly)peptides/proteins.

Legends-to Figures:

Figure 1: General description of the polyphage principle for the display of a Fab library: e.g. library 1: library of VL chains; library 2: VH chains; both libraries on compatible phagemids; in a: libraries are transformed into host cells; in b: library 1 is rescued by a helper phage; in c: libraries are combined by infection; in d: co-expression of heavy and light chains; in e: rescue by helper phages, production of phage particles, assembly of Fab on phage, selection for target; note 1: A certain fraction of the phage particles will be normal unit-length particles containing just one of the two genomes (not shown in Figure 1). Furthermore, polyphage does not discriminate which genomes to package. Therefore, the combinations shown in Figure 1 can arise. To select for

correctly packaged genomes, the subsequent steps are required; in f: infect host cells; in g: select for ability to confer resistance to two antibiotics to infected cells; note 2: only phage that satisfy condition according to g) represent polyphage particles which contain the correct combination of heavy and light chain of binding Fabs (Hetero-polyphage). Unit-length phage as well as polyphage carrying two identical genomes will confer only resistance to one antibiotics.

- Figure 2:** Functional map and sequence of phage vector fhag1A
- Figure 3:** Functional map and sequence of phage vector fjun_1B
- Figure 4:** Functional map and sequence of phage vector fpep3_1B-IR3seq
- Figure 5:** Compatibility of various phage and phagemid vectors: co-transformation of different vector pairs and growth in liquid culture (can/amp selection):
 A. fjun_1B-R408-IR/pIG10_pep10; B. fjun_1B/pIG10_pep10 (only 1 colonie);
 C. fpep3_1B-IR3/pIG10_pep10; D. fjun_1B-R408-IR/pOK1Djun; E. fjun_1B/pOK1Djun: no growth; F. fpep3_1B-IR3/pOK1Djun;
 a. fjun_1B; b. fjun_1B-R408-IR; c. fpep3_1B-IR3; d. pIG10_pep10; e. pOK1Djun
- Figure 6:** co-transformation of positive (pep3/p75ICD combination, lane 9) and negative (jun/p75ICD, lane 10) pairs; lane 1 to 8: SIP transductants
- Figure 7:** Sensitivity of SIP hetero-polyphage system for selection in solution: #SIP hetero-polyphage transductants, transducing units (t.u.)/ml, produced by co-cultures of co-transformants as in Figure 6 mixed at the indicated ratios.
- Figure 8:** PCR to identify phage vector(s) present in SIP polyphage transductants: lane 1 to 6: SIP polyphage transductants; lane A: fpep3_1B-IR3/pIG10.3-IMPp75 co-transformant; lane B: fjun_1B-IR3/pIG10.3-IMPp75 co-transformant
- Figure 9:** IR Phage and Phagemid are Co-packaged into Polyphages: 1: Δ gIII phage + gIII plasmid; 2: IR phage+ phagemid
- Figure 10:** SIP Information is Co-transduced by Polyphages: a: IMPp75 on phage vector; b: pep10-gIII-CT fusion on phage vector; c: IMPp75 on phagemid vector; d: pep10-gIII-CT fusion on phagemid vector

The examples illustrate the invention

Example 1: Selection for polyphage transductants

In WO92/01047, page 83, a model experiment for a two-vector system is described which uses a phage vector (fd-CAT2-IV) encoding a light chain and a phagemid vector (pHEN1-III) encoding a heavy chain. The phagemid, grown in *E. coli* HB2151, was rescued with fd-CAT2-IV phage, and functional phage(mid)s produced. By infecting TG1 cells and plating on tetracycline (to select for fd-CAT) and ampicillin (to select for pHEN1), the ratio of phage and phagemid being packaged was determined.

By repeating this experiment, but plating on TYE plates with both antibiotics, polyphage transductants transducing both resistances simultaneously can be selected, and the genetic information contained on the phage and phagemid vector can be retrieved.

By replacing the single light and heavy chain in the constructs mentioned above by corresponding repertoires, a library of Fab-displaying phage particles can be produced. By screening that library against an immobilized target, a collection of phage particles can be identified. Polyphage particles contained in that collection can be identified by transducing both resistances as described above.

Example 2: Generation and use of an interference-resistant filamentous phage to co-package the genetic information of co-displayed interacting proteins

Introduction

The physical connection of randomly combined genetic information is of vital importance in processes such as interactive screening of two libraries of expressed protein members or for co-expression and co-display of protein pairs which are dependent on the interaction with each other for proper function.

2.1.: Construction of a interference resistant filamentous phage:

2.1.1.: Construction of fjun_1B:

- fflag1A (see Figure 2)

- a. The phage vector f17/9-hag (Krebber *et al.*, 1995, *FEBS Letters* 377, 227-231) is digested with EcoRV and XmnI. The 1.1 kb fragment containing the anti-HAG Ab gene is isolated

by agarose gel electrophoresis and purified with a Qiagen gel extraction kit. This fragment is ligated into a pre-digested pIG10.3 vector (EcoRV-XmnI). Ligated DNA is transformed into DH5a cells and positive clones are verified by restriction analysis. The recombinant clone is called **pIGHag1A**. All cloning described above and subsequently are according to standard protocols (Sambrook *et al.*, 1989, *Molecular Cloning: a Laboratory Manual*, 2nd ed.)

- b. The vector f17/9-hag (Krebber *et al.*, 1995) is digested with EcoRV and StuI. The 7.9 kb fragment is isolated and self-ligated to form the vector **fhag2**.
- c. The chloramphenicol resistance gene (CAT) assembled *via* assembly PCR (Ge and Rudolph, *BioTechniques* 22 (1997) 28-29) using the template pACYC (Cardoso and Schwarz, *J. Appl. Bacteriol.* 72 (1992) 289-293) is amplified by the polymerase chain reaction (PCR) with the primers:
CAT_BspEI(for): 5' GAATGCTCATCCGGAGTTC
CAT_Bsu36I(rev): 5' TTTCAGTGGCCTCAGGCTAGCACCAAGGCGTTTAAG
- d. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with BspEI and Bsu36I then ligated into pre-digested fhag2 vector (BspEI-Bsu36I; 7.2 kb fragment) to form **fhag2C**.
- e. The vector fhag2C is digested with EcoRI and the ends made blunt by filling-in with Klenow fragment. The flushed vector is self-ligated to form vector **fhag2CdelEcoRI**.
- f. pIGHag1A is digested with XbaI and HindIII. The 1.3 kb fragment containing the anti-HAG gene fused with the C-terminal domain of filamentous phage pIII protein is isolated and ligated with a pre-digested fhag2CdelEcoRI phage vector (XbaI-HindIII; 6.4 kb) to create the vector **fhag1A**.

- fjun_1B (see Figure 3)

- a. The DNA encoding the C-terminal domain including the long linker separating it from the amino terminal domain of the filamentous phage pIII (gIII short) is amplified by PCR using pOK1 (Gramatikoff *et al.*, *Nucleic Acids Res.* 22 (1994) 5761-5762) as template with the primers:
gIII short(for): 5'GCTTCCGGAGAATTCAATGCTGGCGGCGGCTCT3'
gIII short(rev): 5'CCCCCCCCAAGCTTATCAAGACTCCTTATTACG3'
- b. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with EcoRI and HindIII, then ligated into pre-digested fhag1A vector (EcoRI-HindIII) to form the vector **fjun_1B**.

2.1.2.: Construction of fjun_1B-R408IR:

In order to introduce mutations which have been described to confer an interference resistance phenotype (Enea and Zinder, Virology 122 (1982), 222-226) into the non-interference resistant fd phage vector fjun_1B (see Fig.3), a 1.7 kb fragment of helper phage R408 (Stratagene) comprising the region between the unique restriction sites *DraIII* and *BsrGI* was PCR amplified by assembly PCR. Subfragments of the 1.7 kb *DraIII*/*BsrGI* fragment were amplified from the f1 phage R408 template DNA with primer combinations FR604/FR605 and FR606/FR607 to introduce via the partially complementary primers FR605 and FR606 an additional *gII* mutation found to be present in the recipient construct fjun_1B. Resulting PCR fragments were gel-purified and combined to serve as template in an subsequent assembly PCR with primers FR604 and FR607. PCR conditions were standard, with approx. 25 ng template, 10 pmole of each primer, 250 pmole of each dNTP, 2 mM Mg, 2.5 U Pfu DNA polymerase (Stratagene). Amplification was done for 30 cycles, with 1 min denaturation at 94 C, 1 min annealing at 50°C, 1 min extension at 72°C. The correct-sized 1.7 kb assembly PCR product was gel-purified, digested with *DraIII* and *BsrGI* and cloned into *DraIII*/*BsrGI*-digested fjun_1B, generating fjun_1B-R408IR.

Primers: FR604 5' GTTCACGTAGTGGGCCATCG 3'
 FR605 5' TGAGAGGTCTAAAAAGGCTATCAGG 3'
 FR606 5' TAGCCTTTTGTAGACCTCTCAAAAATAG 3'
 FR607 5' CGGTGTACAGACCAGGCGC 3'

2.2.: Proof of principle experiments

Despite of the absence of the two originally associated IR mutations, the hybrid phage vector fjun_1B-R408IR (carrying the chloramphenicol acetyltransferase conferring chloramphenicol resistance) could be co-transformed with a phagemid (pOK1deltajun, carrying the beta-lactamase gene conferring ampicillin resistance) containing a phage origin of replication. More importantly, fjun_1B-R408IR could stably co-exist with the phagemid pOK1deltajun, and the phagemid was efficiently co-packaged together with the fjun_1B-R408IR phage genome into polyphage particles. Titers of polyphages, simultaneously

transducing chloramphenicol and ampicillin resistance, reached 6×10^8 transducing units (t.u.)/ml of overnight bacterial culture K91 plating cells, a number almost equivalent to a titer of 10^9 /ml seen after selection on chloramphenicol only. Selection of the K91 transductants on ampicillin only gave a titer of 5×10^9 /ml. These titers indicated that more than 50 % of all phages containing fjun_1B-R408IR also contained the phagemid pOK1deltajun, thus representing polyphages. This high ratio of polyphages was confirmed by restriction analysis of transductants which had been selected on chloramphenicol only. More than 50 % of these clones also contained the phagemid in addition to the fjun_1B-R408IR phage genome. fjun_1B-R408IR was isolated in pure form from an individual transductant, which contained only this phage. The construct fjun_1B-R408IR was used with pOK1deltajun for co-transformation of DH5 α cells, in order to produce selectively-infective phages (SIP) via fos-jun leucine zipper interaction (which non-covalently restores wt gIII function). Stable, double-resistant co-transformants were obtained with this combination and individual clones were grown overnight in the presence of cam/amp. The culture supernatant of these clones was filtered through a 45 μ M membrane filter and used to infect exponentially-growing F+ bacteria (K91 strain) for 20 min at 37 C. To test for the presence of infective SIP polyphages the cells were plated on LB agar plates containing cam and amp and plates were incubated at 37 C overnight. Approx. 500 to 1000 transforming units (t.u.)/ml resulting in double-resistant transductants were obtained from individual co-transformants. DNA of those transductants was analyzed by restriction analysis which showed that 95 % (15/16 clones) of the clones had the correct pattern expected for fjun_1B-R408IR and pOK1deltajun. Supernatants of several polyphage transductants were tested for persistent SIP phage production by re-infection of K91 cells. This confirmed that polyphage transductants continued to produce infective SIP phages and restriction analysis of the resulting 2nd round polyphage transductants showed that 44 % (14/32 clones) contained the correct vector combination. The rest of the clones contained the correct pOK1deltajun phagemid plus a recombined phage vector with a restored wt gIII, indicating an increase in recombination frequency when both vectors are propagated in the rec+ strain K91 (compared to the rec- strain DH5 α used for co-transformation of IR phage and phagemid). To test other protein-protein interactions which give a higher titer of infective SIP phages and to verify the presence of hetero-polyphages (co-packaging of phage and phagemid instead of co-infection by monophages or homo-polyphages), two peptide ligands (previously selected by SIP, WO97/32017)

which bind to the p75 rat neurotrophin receptor (Chao et al., Science 232 (1986) 518-521) intracellular domain (p75ICD) were cloned as N-terminal gIIIc fusions in fjun_1B-R408IR (replacing jun) and the phagemid pIG10.3, leading to constructs fpep3_1B-IR3seq and pIG10.3-pep10 (WO97/32017), respectively, which contain the peptide pep3: 5'-TGTATTGTTTATCATGCTCATTATCTTGTTGCTAAGTGT-3' encoding the amino acid sequence (CysIleValTyrHisAlaHisTyrLeuValAlaLysCys) instead of the jun sequence. Sequencing of the respective parts of the transferred R408 fragment in fpep3_1B-IR3seq revealed that neither of the two IR mutations (the G5986>A mutation from complementation group I in the gII 5' non-translated region, which should be found at position 3225 in fpep3_1B-IR3seq, and the C143>T mutation (3789 in fpep3_1B-IR3seq) from complementation group II leading to a Thr>Ile amino acid exchange in gII) were found to be present. However, the gII mutation G6090>T (3329 in fpep3_1B-IR3seq), leading to a Leu>Val exchange, introduced by assembly PCR was present. Furthermore, three additional mutations compared to an f1 phage could be identified: G5737>A (2976 in fpep3_1B-IR3seq) in the phage origin of replication, G343>A (3989) in gII, and G601>T (4247) in gII/X.

The functional map and the sequence of fpep3_1B-IR3seq are given in Figure 4. This sequence was double-checked several times. It could be shown that differences in the sequence of fpep3_1B-IR3seq compared to published sequence data could be explained by mutations already present in the starting constructs used for cloning fjun_1B-R408IR and fpep3_1B-IR3seq.

Co-transformation experiments (Fig. 5) using combinations of pIG10.3 or pOK1 phagemids (both with f1 oris) with fjun_1B ("wt" fd phage), fjun_1B-R408-IR (containing the DraIII/BsrGI fragment from R408) or fpep3_1B-IR3 (containing the DraIII/BsrGI fragment from R408 and the PCR mutation) revealed that the PCR mutation is not necessary for the IR phenotype, at least judged by the ability to be co-transformable with a phagemid and the ability of individual co-transformants to grow in liquid culture (cam/amp selection).

Additionally, the interacting protein partner p75ICD was cloned as a C-terminal fusion to the infectivity-mediating domains (N1-N2) of gIII (infectivity-mediating particle (IMP) fusion) resulting in constructs fIMPp75-IR3 and pIG10.3-IMPp75.

The IR phage was tested with the SIP pairing fpep3_1B-IR3seq3/ pIG10.3-IMPp75 (which gives a higher titer than fos/jun SIP) in the presence of the negative control combination fjun_1B-IR3seq3/ pIG10.3-IMPp75 (Fig. 6). A SIP hetero-polyphage titer of 1.5×10^5 /ml (cam/amp-resistant transductants) was achieved with fpep3_1B-IR3seq3/ pIG10.3-IMPp75. To test SIP sensitivity in a model library vs. library setting, co-transformants of fpep3_1B-IR3seq3/ pIG10.3-IMPp75 were diluted in an excess fjun_1B-IR3/ pIG10.3-IMPp75 and the supernatant of the bacterial co-culture was assayed for SIP hetero-polyphages. This showed that down to a dilution of 10^{-5} to 10^{-6} can be recovered (Fig. 7).

To prove that only the correct phage vector is present in SIP polyphage transductants, DNA of positive (fpep3_1B-IR3seq3/ pIG10.3-IMPp75) and negative (fjun_1B-IR3/ pIG10.3-IMPp75) control co-transformants, as well as DNA from the SIP polyphage transductants derived from SIP phages produced by the mix of positive and negative control bacteria was analyzed by PCR (Fig. 8). Primers FR614 (5'-GCTCTAGATAACGAGGGC-3') and FR627 (5'-CGCAAGCTTAAGACTCCT-TATTACGC-3') amplify the phage region from the start of ompA to the end of gIII. PCR products derived from fpep3_1B-IR3seq3 and fjun_1B-IR3 can be discriminated by size. Gel analysis of the above samples verified that only the expected fpep3_1B-IR3seq3 phage was present in SIP polyphage transductants (6 analyzed).

To physically demonstrate the existence of hetero-polyphages (which have phage and phagemid co-packaged) when using the IR phage vector, phages produced by co-transformants of fIR3/pIG10.3-IMPp75 and as a control fjun_1B/JB61 ("wt" phage plus complementing gIII plasmid) were separated on an agarose gel (Fig. 9). This showed that the fIR3/pIG10.3-IMPp75 combination produced substantially more slower migrating (thus bigger) phages than the fjun_1B/JB61 control combination. The ratio was almost inversed. Elution of phages from various regions of the gel and subsequent titring of the eluate on plating cells showed that the upper gel region contained a significant portion of double resistance-transducing phages which thus can be regarded as hetero-polyphages.

The pairs fpep3_1B-IR3 and pIG10.3-IMPp75 as well as fIMPp75-IR3 and pIG10.3-pep10 were co-transformed into DH5 α , individual cam/amp resistant clones were grown and the culture supernatant was tested on K91 cells for SIP phage production (Fig. 10). The combinations fpep3_1B-IR3/pIG10.3-IMPp75 and fIMPp75-IR3/pIG10.3-pep10 gave a titer of 1.5×10^5 t.u./ml and 5×10^3 t.u./ml, respectively when assayed for cam/amp-resistant transductants. The titer for each combination when assayed on LB cam was nearly the same as when assayed on LB cam/amp. This demonstrated efficient co-packaging of phage and phagemid DNA to almost 100 %, as seen before with the initial fjun_1B-R408IR and pOK1deltajun combination. To proof the existence of polyphages which individually co-transduce phage and phagemid DNA simultaneously, and to rule out the possibility of transduction of the two resistance markers by independent (and thus random) co-infection by two different phages which have only phage or phagemid packaged, a statistical test was performed. Defined, identical aliquots of bacterial culture supernatants of an individual co-transformant representing each of the two SIP vector combinations described above (fpep3_1B-IR3/pIG10.3-IMPp75 and fIMPp75-IR3/pIG10.3-pep10) were either used individually to infect K91 cells followed by selection on LB cam and LB amp plates, or the same supernatant aliquots from the two vector combinations were mixed before infection of K91 cells and selection on LB cam/amp. 117 cam-resistant, 328 amp-resistant and 141 cam/amp-resistant transforming units were present in the supernatant aliquot from the fIMPp75-IR3/pIG10.3-pep10 combination and 40 cam-resistant, 30 amp-resistant and 23 cam/amp-resistant transforming units were present in the supernatant aliquot from the fpep3_1B-IR3/pIG10.3-IMPp75 combination. The mix of both supernatant aliquots contained 166 cam-resistant and 162 cam/amp-resistant transforming units, exactly corresponding to the expected numbers which would be obtained by adding up the transducing units of the two individual aliquots. 48 cam/amp-resistant transductant colonies were picked from the plate where the mix of the two individual aliquots was used for infection and were analyzed by restriction digest. This showed that only the correct, SIP phage-producing vector combination (5 clones containing the fpep3_1B-IR3/pIG10.3-IMPp75 and 43 clones containing the fIMPp75-IR3/pIG10.3-pep10 combination; this represents a ratio of the two input vector combinations in the analyzed transductants of 1 : 8.6 (fpep3_1B-IR3/pIG10.3-IMPp75 : fIMPp75-IR3/pIG10.3-pep10), which is very similar to the 1 : 6.1 (fpep3_1B-IR3/pIG10.3-IMPp75 : fIMPp75-IR3/pIG10.3-pep10) ratio of double-resistant input phages in this experiment) occurred in all analyzed

transductants, verifying the presence of hetero-polyphages by ruling out the possibility of random co-infection and thus incorrect, random combination by two out of four possible monophage and/or homo-polyphage populations (fpep3_1B-IR3, pIG10.3-IMPp75, fIMPp75-IR3 and pIG10.3-pep10) each containing only one type of vector (phage or phagemid). Statistically, co-infection of the same bacterium by two separate phages was practically already excluded by the small numbers of infective phages containing at least one resistance marker (166 cam-resistant and 358 amp-resistant phages) which were used in the above experiment. Co-infection of the same bacterium (of a total of 10^7 bacteria) by one of the 166 cam-resistant phages and one of the 358 amp-resistant phages has a probability of 6×10^{-10} . Moreover, in this scenario incorrect combinations of individual phage and phagemid vectors (e.g. fpep3_1B-IR3/ pIG10.3-pep10 and fIMPp75-IR3/ pIG10.3-IMPp75) would be possible. The fact that only the correct vector combinations were found in all 48 transductants analyzed from this experiment further proved that co-transduction by hetero-polyphage and not random co-infection by homo-polyphage or monophage was the mechanism by which double-resistance was transduced.

2.3.: Construction of a phage-display system for Fab display

The constructs described in 3.2. can easily be modified to achieve the display of Fabs or a Fab library. In fpep3_1B-IR3seq, the jun part can be replaced by a VL-CL light chain repertoire having the appropriate 3'- and 5'-restriction sites similarly as described for pep_3- to construct fVL_1B-R408IR. In pIG10.3-IMPp75, the IMPp75 construct can be replaced by a repertoire of VH-CH1 heavy chains. After co-transformation of both repertoires into host cells and expression, a library of phage particles displaying Fab fragments is produced. Since fpep3_1B-IR3seq was set up for a SIP experiment by having just the C-terminal domain of gIII, the corresponding Fab-displaying phage particles are non-infectious. By adding a target molecule fused to an infectivity-mediating particle (N1-N2 domain of gIIp), phages displaying target-binding Fab fragments can be selected by infecting host cells.

By replacing the truncated gIII part described above by a full-length copy of gIII, a Fab-display library of infectious phage particles is obtained, which can be screened against immobilized targets. Binding phages can be eluted and used to infect host cells.

By selecting for transductants conferring cam/amp-resistance to their host cells, polyphage infections can be selected in both cases. Thereby the information about both chains of the selected Fab fragments can be retrieved.

CLAIMS

1. A method for identifying a combination of nucleic acid sequences encoding two members of a multimeric (poly)peptide complex with a predetermined property, said combination being contained in a combinatorial library of phage particles displaying a multitude of multimeric (poly)peptides complexes,
said method being characterized by screening or selecting for polyphage particles that contain said combination.
2. The method of claim 1, comprising the steps of
 - (a) providing a first library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences, each encoding a fusion protein of a first member of a multimeric (poly)peptide complex fused to at least part of a phage coat protein, said fusion protein thereby being able to be directed to, and displayed at, the phage surface, wherein said vector molecules are able to be packaged in a phage particle and carry or encode a first selectable and/or screenable property;
 - (b) providing a second library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences, each encoding a second member of a multimeric (poly)peptide complex, wherein the vector molecules of said second library are able to be packaged in a phage particle and carry or encode a second selectable and/or screenable property different from said first property;
 - (c) optionally, providing nucleic acid sequences encoding further members of a multimeric (poly)peptide complex;
 - (d) expressing members of said libraries of recombinant vectors mentioned in steps (a), (b), and optionally nucleic acid sequences mentioned in step (c), in appropriate host cells under appropriate conditions, so that a combinatorial library of phage particles each displaying a multimeric (poly)peptide complex is produced;
 - (e) identifying in said library of phage particles a collection of phages displaying multimeric (poly)peptide complexes having said predetermined property;
 - (f) identifying in said collection polyphage particles simultaneously containing recombinant vector molecules encoding a first and a second member of said

- multimeric (poly)peptide complex by screening or selecting for the simultaneous presence or generation of said first and second selectable and/or screenable property;
- (g) optionally, carrying out further screening and/or selection steps or repeating steps (a) to (f);
- (h) identifying said combination of nucleic acid sequences.

3. The method of claim 1, comprising the steps of

- (a) expressing in appropriate host cells under appropriate conditions
- (aa) genetically diverse nucleic acid sequences contained in a first library of recombinant vector molecules, said nucleic acid sequences comprising a variety of nucleic acid sequences, each encoding a fusion protein of a first member of a multimeric (poly)peptide complex fused to at least part of a phage coat protein, said fusion protein thereby being able to be directed to and displayed at the phage surface, wherein said vector molecules are able to be packaged in a phage particle and carry or encode a first selectable and/or screenable property;
- (ab) genetically diverse nucleic acid sequences contained in a second library of recombinant vector molecules, said nucleic acid sequences comprising a variety of nucleic acid sequences, each encoding a second member of a multimeric (poly)peptide complex, wherein the vector molecules are able to be packaged in a phage particle and carry or encode a second selectable and/or screenable property different from said first property;
- (ac) optionally, nucleic acid sequences encoding further members of a multimeric (poly)peptide complex,
- so that a combinatorial library of phage particles each displaying a multimeric (poly)peptide complex is produced;
- (b) identifying in said library of phage particles a collection of phages displaying multimeric (poly)peptide complexes having said predetermined property;
- (c) identifying in said collection polyphage particles simultaneously containing recombinant vector molecules encoding a first and a second member of said multimeric (poly)peptide complex by screening or selecting for the simultaneous presence or generation of said first and second selectable and/or screenable property;

- (d) optionally, carrying out further screening and/or selection steps or repeating steps (a) to (c);
- (e) identifying said combination of nucleic acid sequences.
4. The method of anyone of claims 1 to 3, wherein the vectors of said first and said second library are a combination of a phage vector and a phagemid vector.
 5. The method of anyone of claims 1 to 3, wherein the vectors of said first and said second library are a combination of two phagemid vectors, said appropriate conditions comprising complementation of phage genes by a helper phage.
 6. The method of claim 5, wherein said two phagemid vectors are compatible.
 7. The method of claim 6, wherein said two phagemid vectors comprise a ColE1 and a p15A plasmid origin of replication.
 8. The method of claim 6, wherein said two phagemid vectors comprise a ColE1 and a mutated ColE1 origin.
 9. The method of anyone of claims 4 to 8, wherein said vectors and/or said helper phage comprise different phage origins of replication.
 10. The method of anyone of claim 4 to 9, wherein said phage vector, said phagemid vector(s) and/or said helper phage are interference resistant.
 11. The method of claim 10, wherein said phage vector, said phagemid vector(s) and/or said helper phage have mutations in the phage intergenic region(s), preferably in positions corresponding to position 5986 of f1, and/or in gene II, preferably in positions corresponding to position 143 of f1.
 12. The method of anyone of claims 10 to 11, wherein said phage vector, said phagemid vector(s) and/or said helper phage are, or are derived from, IR1 mutants such as R176, R382, R383, R407, R408, or from IR2 mutants.

13. The method of anyone of claims 4 to 11, wherein said vectors and/or said helper phage comprise hybrid nucleic acid sequences of f1, fd, and/or M13 derived sequences.
14. The method of anyone of claims 1 to 13, wherein said vector is, or is derived from, fpep3_1B-IR3seq with the sequence listed in Figure 4.
15. The method of claim 14, wherein said derivative is a phage comprising essentially the phage origin or replication from fpep3_1B-IR3seq, the gene II from fpep3_1B-IR3seq, or a combination of said phage origin of replication and said gene II.
16. The method of claim 14, wherein said derivative is a phagemid comprising essentially the phage origin of replication from fpep3_1B-IR3seq, the gene II from fpep3_1B-IR3seq, or a combination of said phage origin of replication and said gene II.
17. The method of claim 14, wherein said derivative is a helper phage comprising essentially the phage origin of replication from fpep3_1B-IR3seq, the gene II from fpep3_1B-IR3seq, or a combination of said phage origin of replication and said gene II.
18. The method of anyone of claims 15 to 17, said derivatives comprise the combined fd/f1 origin including the mutation G5737>A (2976 in fpep3_1B-IR3seq), and/or the mutations G343>A (3989) in gII, and G601>T (4247) in gII/X.
19. The method of anyone of claims 1 to 18, wherein the gene VII contained in any of said vectors contains an amber mutation.
20. The method of claim 19, wherein said mutation is identical to those found in phage vectors R68 or R100.
21. The method of anyone of claims 1 to 20, wherein the gene IX contained in any of said vectors contains an amber mutation.

22. The method of claim 21, wherein said mutation is identical to that found in phage vector N18.
23. The method of anyone of claims 1 to 22, wherein said phage coat protein is gIIIp or gVIIIp.
24. The method of anyone of claims 1 to 23, wherein said phage particles are infectious by having a full-length copy of gIIIp.
25. The method of anyone of claims 1 to 24, wherein said phage particles are non-infectious by having no full-length copy of gIIIp, said fusion protein being formed with a truncated version of gIIIp, wherein the infectivity can be restored by interaction of the displayed multimeric (poly)peptide complexes with a corresponding partner coupled to an infectivity-mediating particle.
26. The method of claim 25, wherein said truncated gIIIp comprises the C-terminal domain of gIIIp.
27. The method of claim 26, wherein said truncated gIIIp is derived from phage fCA55.
28. The method of anyone of claims 1 to 27, wherein said predetermined property is binding to a target.
29. The method of claim 28, wherein said multimeric (poly)peptide complex is a fragment of an immunoglobulin superfamily member.
30. The method of claim 29, wherein said multimeric (poly)peptide complex is a fragment of an immunoglobulin.
31. The method of claim 30, wherein said fragment is an Fv, dsFv or Fab fragment.
32. The method of anyone of claims 1 to 27, wherein said predetermined property is the activity to perform or to catalyze a reaction.

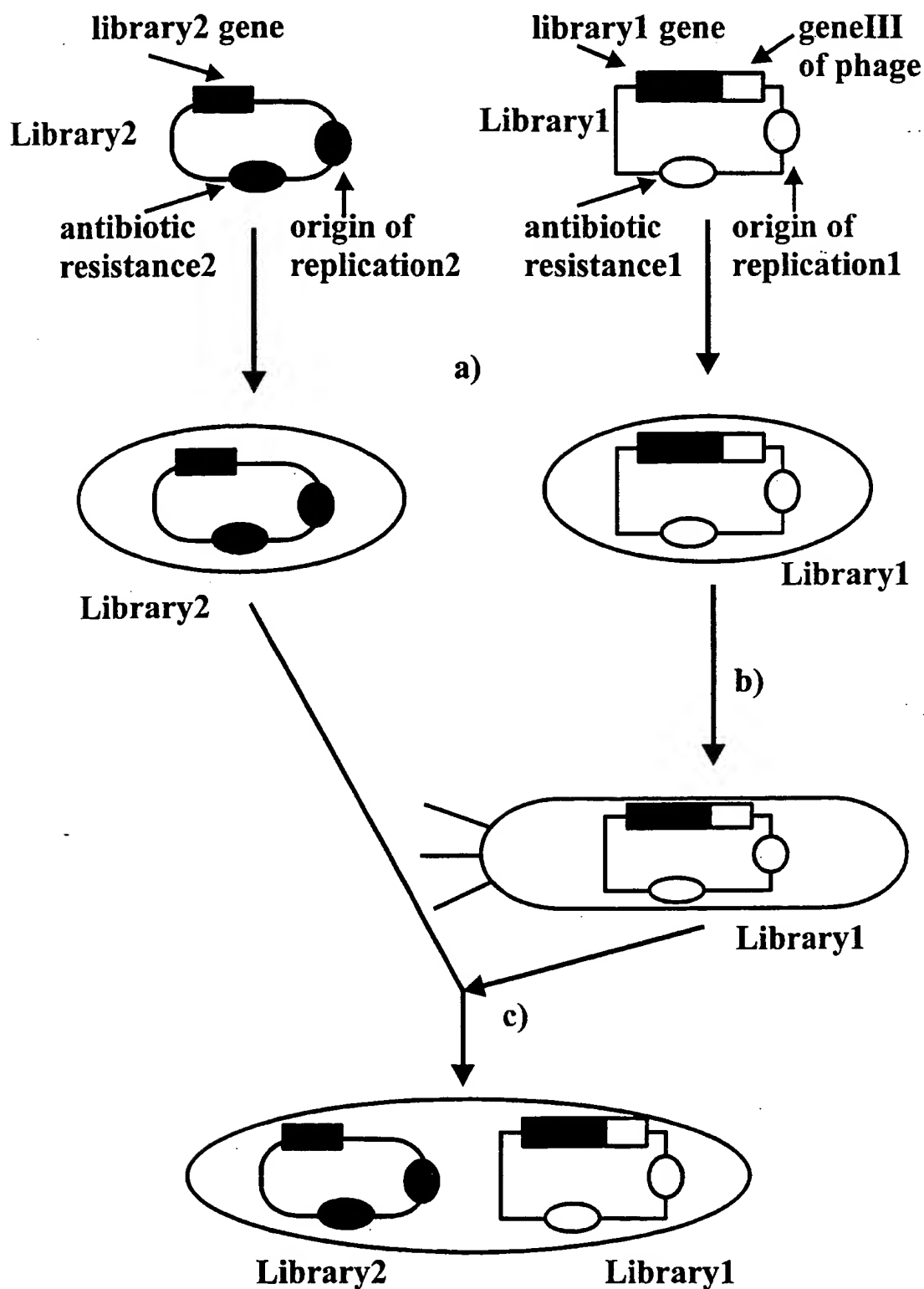
33. The method of claim 32, wherein said multimeric (poly)peptide complex is an enzyme.
34. The method of claim 33, wherein said multimeric (poly)peptide complex is a fragment of a catalytic antibody.
35. The method of claim 34, wherein said fragment is an Fv, dsFv or Fab fragment.
36. The method of anyone of claims 1 to 35, wherein said selectable and/or screenable property is the transactivation of transcription of a reporter gene such as beta-galactosidase, alkaline phosphatase or nutritional markers such as his3 and leu, or resistance genes giving resistance to an antibiotic such as ampicillin, chloramphenicol, kanamycin, zeocin, neomycin, tetracycline or streptomycin.
37. The method of anyone of claims 1 to 36, wherein said generation of said first and second screenable and/or selectable property is achieved after infection of appropriate host cells by said collection of phage particles.
38. The method of anyone of claims 1 to 37, wherein said identification of said nucleic acid sequences is effected by sequencing.
39. The method of anyone of claims 1 to 38, wherein said host cells are E.coli XL-1 Blue, K91 or derivatives thereof, TG1, XL1kann or TOP10F.
40. A polyphage particle which
- (a) contains
 - (i) a first recombinant vector molecule that comprises a nucleic acid sequence, which encodes a fusion protein of a first member of a multimeric (poly)peptide complex fused to at least part of a phage coat protein, and that carries or encodes a first selectable and/or screenable property, and
 - (ii) a second recombinant vector molecule that comprises a nucleic acid sequence, which encodes a second member of a multimeric (poly)peptide complex, and that

carries or encodes a second selectable and/or screenable property different from said first property;
and (b) displays said multimeric (poly)peptide complex at its surface.

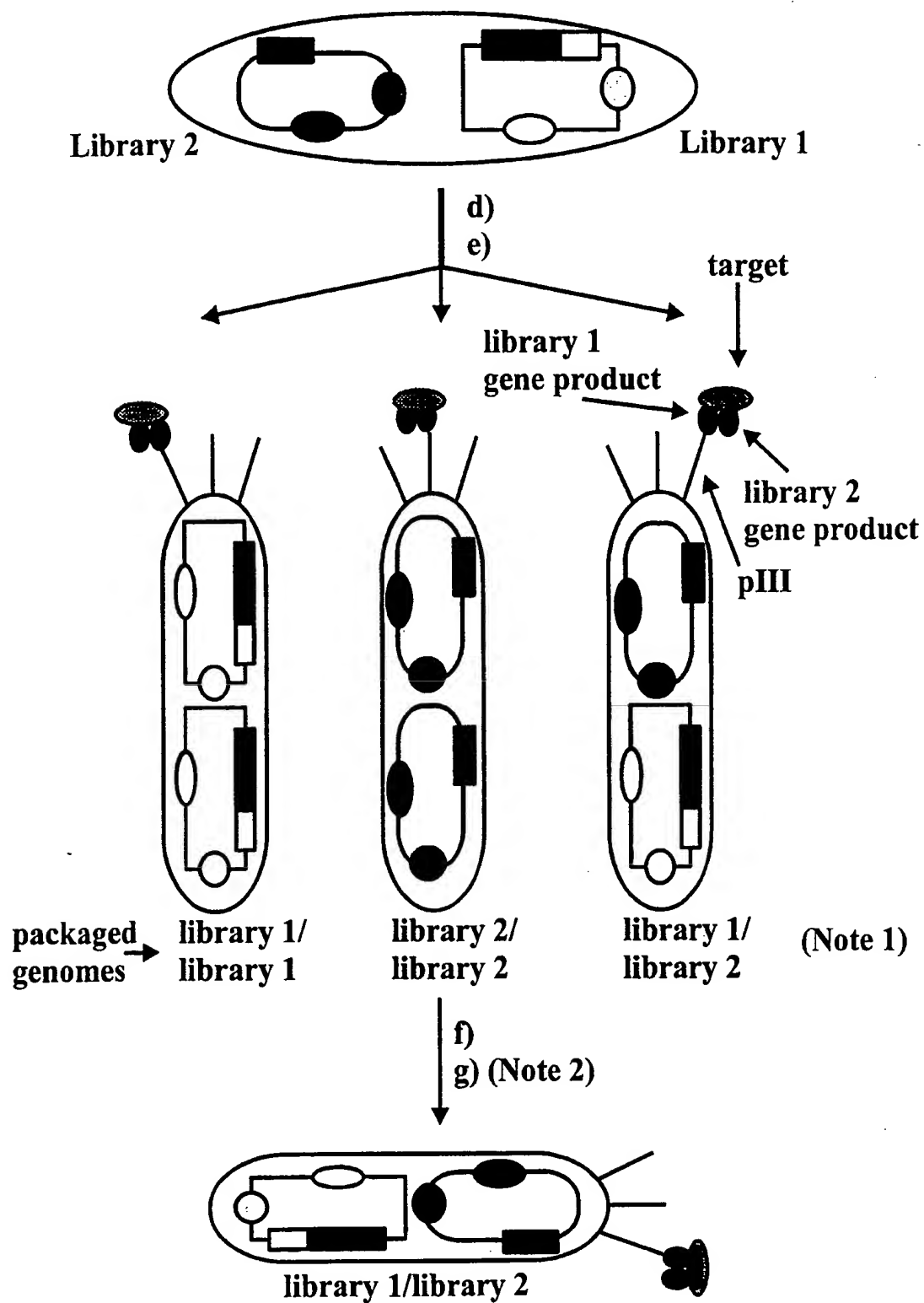
41. The polyphage particle according to claim 40 wherein said phage coat protein is the gIIIp.
42. The polyphage particle according to claim 41 wherein said particles is infectious by having a full-length copy of gIIIp present, either in said fusion protein, or in an additional wild-type copy.
43. The polyphage particle according to claim 41 wherein said particles is non-infectious by having no full-length copy of gIIIp, said fusion protein being formed with a truncated version of gIIIp, wherein the infectivity can be restored by interaction of the displayed multimeric (poly)peptide complex with a corresponding partner coupled to an infectivity-mediating particle.
44. The phage vector fpep3_1B-IR3seq with the sequence listed in Figure 4.
45. A phage vector derived from phage vector fpep3_1B-IR3seq comprising essentially the phage origin or replication from fpep3_1B-IR3seq, the gene II from fpep3_1B-IR3seq, or a combination of said phage origin of replication and said gene II.
46. A phagemid vector derived from phage vector fpep3_1B-IR3seq comprising essentially the phage origin or replication from fpep3_1B-IR3seq, the gene II from fpep3_1B-IR3seq, or a combination of said phage origin of replication and said gene II.
47. A helper phage vector derived from phage vector fpep3_1B-IR3seq comprising essentially the phage origin or replication from fpep3_1B-IR3seq, the gene II from fpep3_1B-IR3seq, or a combination of said phage origin of replication and said gene II.
48. A vector according to anyone of claims 45 to 47, wherein said derivatives comprise the combined fd/fl origin including the mutation G5737>A (2976 in fpep3_1B-IR3seq), and/or the mutations G343>A (3989) in gII, and G601>T (4247) in gII/X.

49. The use according to any of the vectors of anyone of claims 44 to 48 in the generation of polyphage particles containing a combination of at least two different vectors.
50. The use according to claim 49, wherein said combination of different vectors comprises nucleic acid sequences encoding members of a multimeric (poly)peptide complex.
51. The use according to claim 50, wherein said combination of different vectors comprises nucleic acid sequences encoding interacting (poly)peptides/proteins.

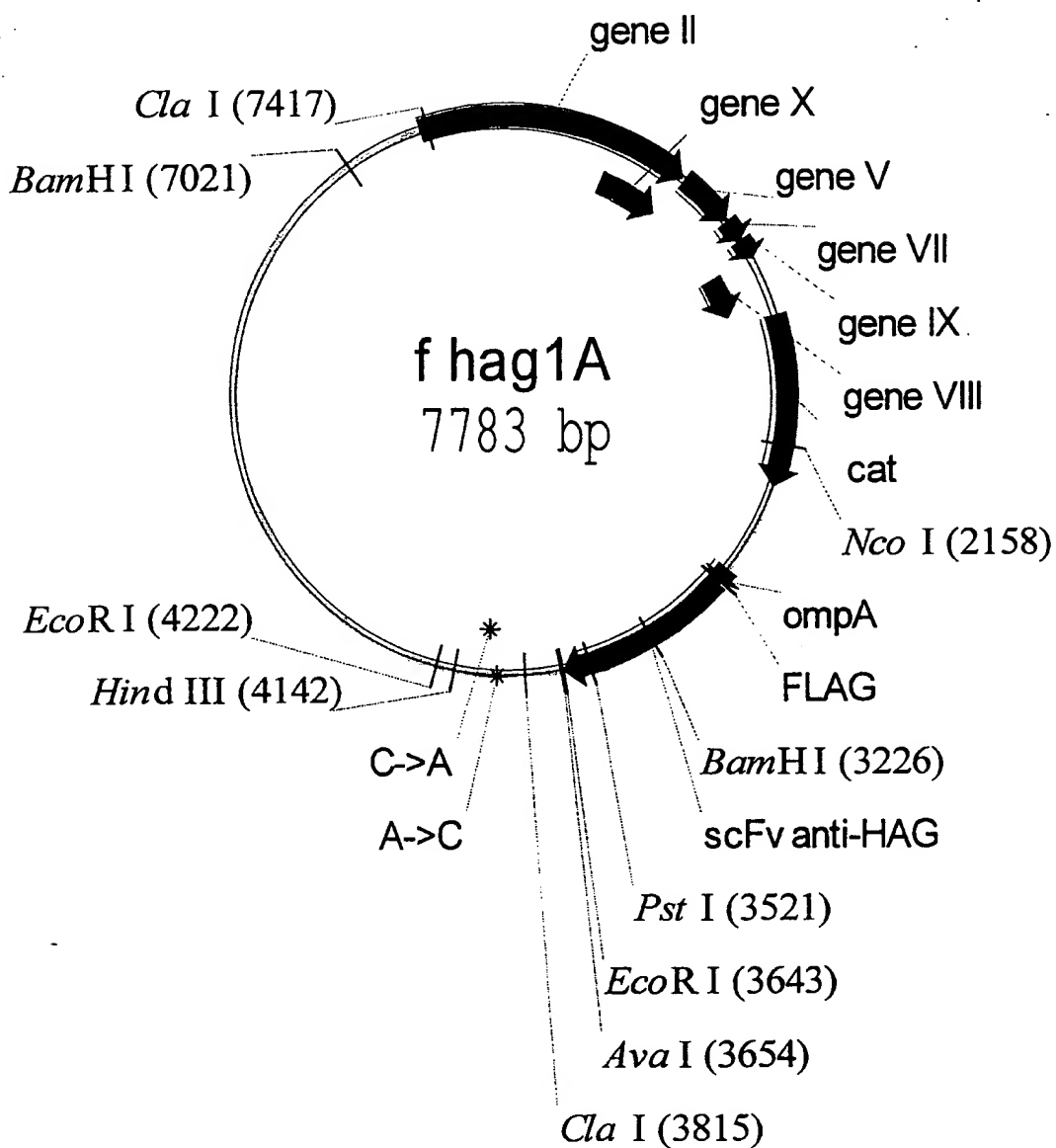
1/39

Figure 1: General description of the polyphage principle

2/39

Figure 1: General description of the polyphage principle (cont.)

3/39

Figure 2

4/39

1	AACGCTACTA	CCATTAGTAG	AATTGATGCC	ACCTTTTCAG	CTCGCGCCCC
	TTGCGATGAT	GGTAATCATC	TTAACTACGG	TGGAAAAGTC	GAGCGCGGGG
51	AAATGAAAAT	ATAGCTAAAC	AGGTTATTGA	CCATTTGCGA	AATGTATCTA
	TTTACTTTTA	TATCGATTTG	TCCAATAACT	GGTAAACGCT	TTACATAGAT
101	ATGGTCAAAC	TAAATCTACT	CGTTCGCAGA	ATTGGGAATC	AACTGTTACA
	TACCAAGTTT	ATTTAGATGA	GCAAGCGTCT	TAACCCTTAG	TTGACAATGT
151	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA	GTTGCATATT	TAAAACATGT
	ACCTTACTTT	GAAGGTCTGT	GGCATGAAAT	CAACGTATAA	ATTTTGTACA
201	TGAACTACAG	CACCAGATTC	AGCAATTAAG	CTCTAAGCCA	TCCGCAAAAA
	ACTTGATGTC	GTGGTCTAAG	TCGTTAATTC	GAGATTCGGT	AGGCGTTTTT
251	TGACCTCTTA	TCAAAAGGAG	CAATTAAAGG	TACTGTCTAA	TCCTGACCTG
	ACTGGAGAAT	AGTTTTCTTC	GTTAATTTCC	ATGACAGATT	AGGACTGGAC
301	TTGGAATTTG	CTTCCGGTCT	GGTTCGCTTT	GAGGCTCGAA	TTGAAACGCG
	AACCTTAAAC	GAAGGCCAGA	CCAAGCGAAA	CTCCGAGCTT	AACCTTTGCGC
351	ATATTTGAAG	TCTTTCGGGC	TTCCTCTTAA	TCTTTTTGAT	GCAATTCGCT
	TATAAACTTC	AGAAAGCCCG	AAGGAGAATT	AGAAAACTA	CGTTAAGCGA
401	TTGCTTCTGA	CTATAATAGA	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG
	AACGAAGACT	GATATTATCT	GTCCCATTTT	TGGACTAAAA	ACTAAATACC
451	TCATTCTCGT	TTTCTGAACT	GTTTAAAGCA	TTTGAGGGGG	ATTCAATGAA
	AGTAAGAGCA	AAAGACTTGA	CAAATTTCTG	AAACTCCCCC	TAAGTTACTT
501	TATTTATGAC	GATTCCGCAG	TATTGGACGC	TATCCAGTCT	AAACATTTTA
	ATAAATACTG	CTAAGGCGTC	ATAACCTGCG	ATAGGTCAGA	TTTGTAAAAAT
551	CAATTACCCC	CTCTGGCAAA	ACTTCCTTTG	CAAAAGCCTC	TCGCTATTTT
	GTTAATGGGG	GAGACCGTTT	TGAAGGAAAC	GTTTTCTGGAG	AGCGATAAAA
601	GGTTTCTATC	GTCGTCTGGT	TAATGAGGGT	TATGATAGTG	TTGCTCTTAC
	CCAAAGATAG	CAGCAGACCA	ATTACTCCCA	ATACTATCAC	AACGAGAATG
651	CATGCCTCGT	AATTCCTTTT	GGCGTTATGT	ATCTGCATTA	GTTGAGTGTG
	GTACGGAGCA	TTAAGGAAAA	CCGCAATACA	TAGACGTAAT	CAACTCACAC
701	GTATTCCTAA	ATCTCAATTG	ATGAATCTTT	CCACCTGTAA	TAATGTTGTT
	CATAAGGATT	TAGAGTTAAC	TACTTAGAAA	GGTGGACATT	ATTACAACAA
751	CCGTTAGTTC	GTTTTATTAA	CGTAGATTTT	TCCTCCCAAC	GTCCTGACTG
	GGCAATCAAG	CAAAATAATT	GCATCTAAAA	AGGAGGGTTG	CAGGACTGAC
801	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA	AAATGATTAA
	CATATTACTC	GGTCAAGAAT	TTTAGCGTAT	TCCATTAAGT	TTTACTAATT

5/39

851	AGTTGAAATT	AAACCGTCTC	AAGCGCAATT	TACTACCCGT	TCTGGTGTTC
	TCAACTTTAA	TTTGGCAGAG	TTCGCGTTAA	ATGATGGGCA	AGACCACAAA
901	CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTTTG	TTACGTTGAT
	GAGCAGTCCC	GTTTCGGAATA	AGTGACTTAC	TCGTCGAAAC	AATGCAACTA
951	TTGGGTAATG	AATATCCGGT	GCTTGTCAAG	ATTACTCTCG	ACGAAGGTCA
	AACCCATTAC	TTATAGGCCA	CGAACAGTTC	TAATGAGAGC	TGCTTCCAGT
1001	GCCAGCGTAT	GCGCCTGGTC	TGTACACCGT	GCATCTGTCC	TCGTTCAAAG
	CGGTCGCATA	CGCGGACCAG	ACATGTGGCA	CGTAGACAGG	AGCAAGTTTC
1051	TTGGTCAGTT	CGGTTCTCTT	ATGATTGACC	GTCTGCGCCT	CGTTCCGGCT
	AACCAGTCAA	GCCAAGAGAA	TACTAACTGG	CAGACGCGGA	GCAAGGCCGA
1101	AAGTAACATG	GAGCAGGTCTG	CGGATTTCTGA	CACAATTTAT	CAGGCGATGA
	TTCATTGTAC	CTCGTCCAGC	GCCTAAAGCT	GTGTTAAATA	GTCCGCTACT
1151	TACAAATCTC	CGTTGTACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGGT
	ATGTTTAGAG	GCAACATGAA	ACAAAGCGCG	AACCATATTA	GCGACCCCCA
1201	CAAAGATGAG	TGTTTTAGTG	TATTCTTTCTG	CCTCTTTCTG	TTTAGGTTGG
	GTTTCTACTC	ACAAAATCAC	ATAAGAAAGC	GGAGAAAGCA	AAATCCAACC
1251	TGCCTTCGTA	GTGGCATTAC	GTATTTTACC	CGTTTAATGG	AAACTTCCTC
	ACGGAAGCAT	CACCGTAATG	CATAAAATGG	GCAAATTACC	TTTGAAGGAG
1301	ATGCGTAAGT	CTTTAGTCCT	CAAAGCCTCC	GTAGCCGTTG	CTACCCCTCGT
	TACGCATTCA	GAAATCAGGA	GTTTCGGAGG	CATCGGCAAC	GATGGGAGCA
1351	TCCGATGCTG	TCTTTCGCTG	CTGAGGGTGA	CGATCCCGCA	AAAGCGGCCT
	AGGCTACGAC	AGAAAGCGAC	GACTCCCACT	GCTAGGGCGT	TTTCGCCGGA
1401	TTGACTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA	TGCGTGGGCG
	AACTGAGGGA	CGTTCGGAGT	CGCTGGCTTA	TATAGCCAAT	ACGCACCCGC
1451	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA
	TACCAACAAC	AGTAACAGCC	GCGTTGATAG	CCATAGTTCTG	ACAAATTCTT
1501	ATTCACCTCG	AAAGCAAGCT	GATAAAGGAG	GTTTCTCGAT	CGAGACGTTN
	TAAGTGGAGC	TTTCGTTCGA	CTATTTCTCTC	CAAAGAGCTA	GCTCTGCAAN
1551	NNNGAGGTTC	CAACTTTCAC	CATAATGAAA	TAAGATCACT	ACCGGGCGTA
	NNNCTCCAAG	GTTGAAAGTG	GTATTACTTT	ATTCTAGTGA	TGGCCCCCAT
1601	TTTTTTGAGT	TATCGAGATT	TTCAGGAGCT	AAGGAAGCTA	AAATGGAGAA
	AAAAAACTCA	ATAGCTCTAA	AAGTCCTCGA	TTCTTTCGAT	TTTACCTCTT
1651	AAAAATCACT	GGATATACCA	CCGTTGATAT	ATCCCAATGG	CATCGTAAAG
	TTTTTAGTGA	CCTATATGGT	GGCAACTATA	TAGGGTTACC	GTAGCATTTT

6/39

1701	AACATTTTGA	GGCATTTCAG	TCAGTTGCTC	AATGTACCTA	TAACCAGACC
	TTGTAAAAC	CCGTAAAGTC	AGTCAACGAG	TTACATGGAT	ATTGGTCTGG
1751	G TTCAGCTGG	ATATTACGGC	CTTTTTTAAAG	ACCGTAAAGA	AAAATAAGCA
	CAAGTCGACC	TATAATGCCG	GAAAAATTTC	TGGCATTCT	TTTTATTCTG
1801	CAAGTTTTAT	CCGGCCTTTA	TTCACATTCT	TGCCCCCCTG	ATGAATGCTC
	GTTCAAAAATA	GGCCGGAAT	AAGTGTAAGA	ACGGGCGGAC	TACTTACGAG
1851	ATCCGGAGTT	CCGTATGGCA	ATGAAAGACG	GTGAGCTGGT	GATATGGGAT
	TAGGCCTCAA	GGCATACCGT	TACTTTCTGC	CACTCGACCA	CTATACCCTA
1901	AGTGTTTACC	CTTGTTACAC	CGTTTTCCAT	GAGCAAACTG	AAACGTTTTTC
	TCACAAAGTG	GAACAATGTG	GCAAAAAGGTA	CTCGTTTGAC	TTTGCAAAAG
1951	ATCGCTCTGG	AGTGAATACC	ACGACGATTT	CCGGCAGTTT	CTACACATAT
	TAGCGAGACC	TCACTTATGG	TGCTGCTAAA	GGCCGTCAAA	GATGTGTATA
2001	ATTCGCAAGA	TGTGGCGTGT	TACGGTGAAA	ACCTGGCCTA	TTTCCCTAAA
	TAAGCGTTCT	ACACCGCACA	ATGCCACTTT	TGGACCGGAT	AAAGGGATTT
2051	GGGTTTATTG	AGAATATGTT	TTTCGTCTCA	GCCAATCCCT	GGGTGAGTTT
	CCCAAATAAC	TCTTATACAA	AAAGCAGAGT	CGGTTAGGGA	CCCACTCAAA
2101	CACCAGTTTT	GATTTAAACG	TGGCCAATAT	GGACAACTTC	TTCGCCCCCG
	GTGGTCAAAA	CTAAATTTGC	ACCGTTTATA	CCTGTTGAAG	AAGCGGGGGC

NcoI

~~~~~

|      |             |             |            |            |             |
|------|-------------|-------------|------------|------------|-------------|
| 2151 | TTTTCAACCAT | GGGCAAATAT  | TATACGCAAG | GCGACAAGGT | GCTGATGCCG  |
|      | AAAAGTGGTA  | CCCGTTTATA  | ATATGCGTTC | CGCTGTTCCA | CGACTACGGC  |
| 2201 | CTGGCGATT   | AGGTTTATCA  | TGCCGTCTGT | GATGGCTTCC | ATGTCGGCAG  |
|      | GACCGCTAAG  | TCCAAGTAGT  | ACGGCAGACA | CTACCGAAGG | TACAGCCGTC  |
| 2251 | AATGCTTAAT  | GAATTACAAC  | AGTACTGCGA | TGAGTGGCAG | GGCGGGGCGT  |
|      | TTACGAATTA  | CTTAATGTTG  | TCATGACGCT | ACTCACCGTC | CCGCCCCGCA  |
| 2301 | AATTTTTTTTA | AGGCAGTTAT  | TGGTGCCCTT | AAACGCCTGG | TGCTACGCCT  |
|      | TTAAAAAAAT  | TCCGTCAATA  | ACCACGGGAA | TTTGCGGACC | ACGATGCGGA  |
| 2351 | GAATAAGTGA  | TAATAAGCGG  | ATGAATGGCA | GAAATTCGAA | AGCAAATTTCG |
|      | CTTATTCACT  | ATTATTCGCC  | TACTTACCGT | CTTTAAGCTT | TCGTTTAAGC  |
| 2401 | ACCCGGTCGT  | CGGTTTCAAGG | CAGGGTCGTT | AAATAGCCGC | TTATGTCTAT  |
|      | TGGGCCAGCA  | GCCAAGTCCC  | GTCCAGCAA  | TTTATCGGCG | AATACAGATA  |
| 2451 | TGCTGGTTTA  | CCGGTTTATT  | GACTACCGGA | AGCAGTGTGA | CCGTGTGCTT  |
|      | ACGACCAAAT  | GGCCAAATAA  | CTGATGGCCT | TCGTCACT   | GGCACACGAA  |
| 2501 | CTCAAATGCC  | TGAGGCCAGT  | TTGCTCAGGC | TCTCCCCGTG | GAGGTAATAA  |
|      | GAGTTTACGG  | ACTCCGGTCA  | AACGAGTCCG | AGAGGGGCAC | CTCCATTATT  |

7/39

2551 TTGCTCGACC GATAAAAGCG GCTTCCTGAC AGGAGGCCGT TTTGTTTTGC  
 AACGAGCTGG CTATTTTCGC CGAAGGACTG TCCTCCGGCA AAACAAAACG  
 2601 AGCCACCTC AACGCAATTA ATGTGAGTTA GCTCACTCAT TAGGCACCCC  
 TCGGGTGGAG TTGCGTTAAT TACACTCAAT CGAGTGAGTA ATCCGTGGGG  
 2651 AGGCTTTACA CTTTATGCTT CCGGCTCGTA TGTGTGTGG AATTGTGAGC  
 TCCGAAATGT GAAATACGAA GGCCGAGCAT ACAACACACC TTAACACTCG  
 2701 GGATAACAAT TTCACACAGG AAACAGCTAT GACCATGATT ACGAATTTCT  
 CCTATTGTTA AAGTGTGTCC TTTGTCGATA CTGGTACTAA TGCTTAAAGA  
 2751 AGATAACGAG GGCAAATCAT GAAAAAGACA GCTATCGCGA TTGCAGTGGC  
 TCTATTGCTC CCGTTTAGTA CTTTTTCTGT CGATAGCGCT AACGTCACCG  
 2801 ACTGGCTGGT TTCGCTACCG TAGCGCAGGC CGACTACAAA GATATCGTTA  
 TGACCGACCA AAGCGATGGC ATCGCGTCCG GCTGATGTTT CTATAGCAAT  
 2851 TGACCCAGTC ACCGTCTCTC CTGACCGTTA CCGCTGGTGA AAAAGTTACC  
 ACTGGGTCAG TGGCAGGAGG GACTGGCAAT GGCGACCACT TTTTCAATGG  
 2901 ATGTCCTGCA CCTCCTCCCA GTCCCTGTTT AACTCCGGTA AACAGAAAAA  
 TACAGGACGT GGAGGAGGGT CAGGGACAAG TTGAGGCCAT TTGTCTTTTT  
 2951 CTACCTGACC TGGTATCAGC AGAAACCGGG TCAGCCACCG AAAGTTCTGA  
 GATGGACTGG ACCATAGTCG TCTTTGGCCC AGTCGGTGGC TTTCAAGACT  
 3001 TCTACTGGGC TTCCACCCGT GAATCCGGTG TTCCAGACCG TTTCACCGGT  
 AGATGACCCG AAGGTGGGCA CTTAGGCCAC AAGGTCTGGC AAAGTGGCCA  
 3051 TCCGGTTCCG GCACCGACTT CACCCTGACC ATCTCCTCCG TTCAGGCTGA  
 AGGCCAAGGC CGTGGCTGAA GTGGGACTGG TAGAGGAGGC AAGTCCGACT  
 3101 AGACCTGGCT GTTTACTACT GCCAGAACGA CTACTCCAAC CCACTGACCT  
 TCTGGACCGA CAAATGATGA CGGTCTTGCT GATGAGGTTG GGTGACTGGA  
 3151 TCGGTGGTGG CACCAAACCTG GAACTTAAGC GCGCTGGTGG TGGAGGGTCT  
 AGCCACCACC GTGGTTTGAC CTTGAATTCG CGCGACCACC ACCTCCCAGA

BamHI

~~~~~

3201 GGAGGAGGTG GGAGTGGGGG AGGTGGATCC GGCGGGGGAG GTTCAGGGGG
 CCTCCTCCAC CCTCACCCCC TCCACCTAGG CCGCCCCCTC CAAGTCCCCC
 3251 TGGCGGTAGT GGAGGGGGCG GTTCAGAAGT TCAACTAGTT GAATCCGGTG
 ACCGCCATCA CCTCCCCCGC CAAGTCTTCA AGTTGATCAA CTTAGGCCAC
 3301 GTGACCTGGT TAAACCGGGT GGTTCCTTGA AACTGTCCTG CGCTGCTTCC
 CACTGGACCA ATTTGGCCCA CCAAGGACTT TTGACAGGAC GCGACGAAGG

8/39

3351 GGTTCCTCCT TCTCCTCCTA CGGTATGTCC TGGGTTCGTC AGACCCCGGA
 CCAAAGAGGA AGAGGAGGAT GCCATACAGG ACCCAAGCAG TCTGGGGCCT

3401 CAAACGTCTG GAATGGGTTG CTACCATCTC CAACGGTGGT GGTACACCT
 GTTTGCAGAC CTTACCCAAC GATGGTAGAG GTTGCCACCA CCAATGTGGA

3451 ACTACCCGGA CTCCGTTAAA GGTCGTTTCA CCATCTCCCG TGACAACGCT
 TGATGGGCCT GAGGCAATTT CCAGCAAAGT GGTAGAGGGC ACTGTTGCGA

PstI

~~~~~

3501 AAAAACACCC TGTACCTGCA GATGTCTCCTC CTGAAATCCG AAGACTCAGC  
 TTTTGTGGG ACATGGACGT CTACAGGAGG GACTTTAGGC TTCTGAGTCG

3551 TATGTACTAC TGCCTCGTC GTGAACGTTA CGACGAAAAC GGTTCGCTT  
 ATACATGATG ACGCGAGCAG CACTTGCAAT GCTGCTTTTG CCAAAGCGAA

EcoRI

~~~~~

3601 ACTGGGGTCA GGTACCTG GTTACCGTTT CAGCTTCCGG AGAATTCGAG
 TGACCCAGT CCCATGGGAC CAATGGCAAA GTCGAAGGCC TCTTAAGCTC

AvaI

~~~~~

3651 GCCTCGGGGG CCGAGGGCGG CGGTTCTGGT TCCGGTGATT TTGATTATGA  
 CGGAGCCCCC GGCTCCCGCC GCCAAGACCA AGGCCACTAA AACTAATACT

3701 AAAAATGGCA AACGCTAATA AGGGGGCTAT GACCGAAAAT GCCGATGAAA  
 TTTTACCCTG TTGCGATTAT TCCCCGATA CTGGCTTTTA CGGCTACTTT

3751 ACGCGCTACA GTCTGACGCT AAAGGCAAAC TTGATTCTGT CGCTACTGAT  
 TCGCGGATGT CAGACTGCGA TTTCCGTTTG AACTAAGACA GCGATGACTA

ClaI

~~~~~

3801 TACGGTGCTG CTATCGATGG TTTCATTGGT GACGTTTCCG GCCTTGCTAA
 ATGCCACGAC GATAGCTACC AAAGTAACCA CTGCAAAGGC CGGAACGATT

3851 TGGTAATGGT GCTACTGGTG ATTTTGCTGG CTCTAATTCC CAAATGGCTC
 ACCATTACCA CGATGACCAC TAAACGACC GAGATTAAGG GTTTACCGAG

3901 AAGTCGGTGA CGGTGATAAT TCACCTTTAA TGAATAATTT CCGTCAATAT
 TTCAGCCACT GCCACTATTA AGTGGAATTT ACTTATTAAA GGCAGTTATA

3951 TTACCTTCCC TCCCTCAATC GGTTGAATGT CGCCCTTTTG TCTTTGGCGC
 AATGGAAGGG AGGGAGTTAG CCAACTTACA GCGGGAAC AGAAACCGCG

4001 TGGTAAACCA TATGAATTTT CTATTGATTG TGACAAAATA AACTTATTCC
 ACCATTTGGT ATACTTAAAA GATAACTAAC ACTGTTTTAT TTGAATAAGG

4051 GTGGTGTCTT TGCCTTTCTT TTATATGTTG CCACCTTTAT GTATGTATTT
 CACCACAGAA ACGCAAAGAA AATATACAAC GGTGGAAATA CATACATAAA

9/39

HindIII

4101 TCTACGTTTG CTAACATACT GCGTAATAAG GAGTCTTGAT AAGCTTCGAG
AGATGCAAAC GATTGTATGA CGCATTATTC CTCAGAACTA TTCGAAGCTC

4151 AAATTCACCT CGAAAGCAAG CTGATAAACC GATACAATTA AAGGCTCCTT
TTTAAGTGGA GCTTTCGTTC GACTATTGCG CTATGTAAAT TTCCGAGGAA

EcoRI

~~~~~

4201 TTGGAGCCTT TTTTTTTGGA GAATTCAATC ATGCCAGTTC TTTTGGGTAT  
AACCTCGGAA AAAAAAACCT CTTAAGTTAG TACGGTCAAG AAAACCCATA

4251 TCCGTTATTA TTGCGTTTCC TCGGTTTCCT TCTGGTAACT TTGTTCGGCT  
AGGCAATAAT AACGCAAAGG AGCCAAAGGA AGACCATTGA AACAAGCCGA

4301 ATCTGCTTAC TTCCTTAAA AAGGGCTTCG GTAAGATAGC TATTGCTATT  
TAGACGAATG AAAGGAATTT TTCCCGAAGC CATTCTATCG ATAACGATAA

4351 TCATTGTTTC TTGCTCTTAT TATTGGGCTT AACTCAATTC TTGTGGGTTA  
AGTAACAAAG AACGAGAATA ATAACCCGAA TTGAGTTAAG AACACCCAAT

4401 TCTCTCTGAT ATTAGCGCAC AATTACCCTC TGATTTTGTT CAGGGCGTTC  
AGAGAGACTA TAATCGCGTG TTAATGGGAG ACTAAAACAA GTCCCGCAAG

4451 AGTTAATTCT CCCGTCTAAT GCGCTTCCCT GTTTTTATGT TATTCTCTCT  
TCAATTAAGA GGGCAGATTA CGCGAAGGGA CAAAAATACA ATAAGAGAGA

4501 GTAAAGGCTG CTATTTTTCAT TTTTGACGTT AAACAAAAAA TCGTTTCTTA  
CATTTCCGAC GATAAAAGTA AAAACTGCAA TTTGTTTTTT AGCAAAGAAT

4551 TTTGGATTGG GATAAATAAA TATGGCTGTT TATTTTGTA CTGGCAAATT  
AAACCTAACC CTATTTATTT ATACCGACAA ATAAAACATT GACCGTTTAA

4601 AGGCTCTGGA AAGACGCTCG TTAGCGTTGG TAAGATTCAG GATAAAATTG  
TCCGAGACCT TTCTGCGAGC AATCGCAACC ATTCTAAGTC CTATTTTAAC

4651 TAGCTGGGTG CAAAATAGCA ACTAATCTTG ATTTAAGGCT TCAAAACCTC  
ATCGACCCAC GTTTTATCGT TGATTAGAAC TAAATTCCGA AGTTTGGGAG

4701 CCGCAAGTCG GGAGGTTCCG TAAAACGCCT CGCGTTCTTA GAATACCGGA  
GGCGTTCAGC CCTCCAAGCG ATTTTGCGGA GCGCAAGAAT CTTATGGCCT

4751 TAAGCCTTCT ATTTCTGATT TGCTTGCTAT TGGTCGTGGT AATGATTCCT  
ATTCGGAAGA TAAAGACTAA ACGAACGATA ACCAGCACCA TTAATAAGGA

4801 ACGACGAAAA TAAAAACGGT TTGCTTGTTT TTGATGAATG CGGTACTTGG  
TGCTGCTTTT ATTTTGCCA AACGAACAAG AACTACTTAC GCCATGAACC

4851 TTTAATACCC GTTCATGGAA TGACAAGGAA AGACAGCCGA TTATTGATTG  
AAATTATGGG CAAGTACCTT ACTGTTCCCT TCTGTCGGCT AATAACTAAC

10/39

|      |            |             |             |            |             |
|------|------------|-------------|-------------|------------|-------------|
| 4901 | GTTTCTTCAT | GCTCGTAAAT  | TGGGATGGGA  | TATTATTTTT | CTTGTTTCAGG |
|      | CAAAGAAGTA | CGAGCATTTA  | ACCCTACCCCT | ATAATAAAAA | GAACAAGTCC  |
| 4951 | ATTTATCTAT | TGTTGATAAA  | CAGGCGCGTT  | CTGCATTAGC | TGAACACGTT  |
|      | TAAATAGATA | ACAAC TATTT | GTCCGCGCAA  | GACGTAATCG | ACTTGTGCAA  |
| 5001 | GTTTATTGTC | GCCGTCTGGA  | CAGAATTACT  | TTACCCTTTG | TCGGCACTTT  |
|      | CAAATAACAG | CGGCAGACCT  | GTCTTAATGA  | AATGGGAAAC | AGCCGTGAAA  |
| 5051 | ATATTCTCTT | GTTACTGGCT  | CAAAAATGCC  | TCTGCCTAAA | TTACATGTTG  |
|      | TATAAGAGAA | CAATGACCGA  | GTTTTTACGG  | AGACGGATTT | AATGTACAAC  |
| 5101 | GTGTTGTAA  | ATATGGTGAT  | TCTCAATTAA  | GCCCTACTGT | TGAGCGTTGG  |
|      | CACAACAATT | TATACCACTA  | AGAGTTAATT  | CGGGATGACA | ACTCGCAACC  |
| 5151 | CTTTATACTG | GTAAGAATTT  | ATATAACGCA  | TATGACACTA | AACAGGCTTT  |
|      | GAAATATGAC | CATTCTTAAA  | TATATTGCGT  | ATACTGTGAT | TTGTCCGAAA  |
| 5201 | TTCCAGTAAT | TATGATT CAG | GTGTTTATTC  | ATATTTAACC | CCTTATTTAT  |
|      | AAGGTCATTA | ATACTAAGTC  | CACAAATAAG  | TATAAATTGG | GGAATAAATA  |
| 5251 | CACACGGTCG | GTATTTCAAA  | CCATTAAATT  | TAGGTCAGAA | GATGAAATTA  |
|      | GTGTGCCAGC | CATAAAGTTT  | GGTAATTTAA  | ATCCAGTCTT | CTACTTTAAT  |
| 5301 | ACTAAAATAT | ATTTGAAAAA  | GTTTTCTCGC  | GTTCTTTGTC | TTGCGATAGG  |
|      | TGATTTTATA | TAAACTTTTT  | CAAAAGAGCG  | CAAGAAACAG | AACGCTATCC  |
| 5351 | ATTTGCATCA | GCATTTACAT  | ATAGTTATAT  | AACCCAACCT | AAGCCGGAGG  |
|      | TAAACGTAGT | CGTAAATGTA  | TATCAATATA  | TTGGGTTGGA | TTCGGCCTCC  |
| 5401 | TTAAAAAGGT | AGTCTCTCAG  | ACCTATGATT  | TTGATAAATT | CACTATTGAC  |
|      | AATTTTTCCA | TCAGAGAGTC  | TGGATACTAA  | AACTATTTAA | GTGATAACTG  |
| 5451 | TCTTCTCAGC | GTCTTAATCT  | AAGCTATCGC  | TATGTTTTCA | AGGATTCTAA  |
|      | AGAAGAGTCG | CAGAATTAGA  | TTCGATAGCG  | ATACAAAAGT | TCCTAAGATT  |
| 5501 | GGGAAAATTA | ATTAATAGCG  | ACGATTTACA  | GAAGCAAGGT | TATTCCATCA  |
|      | CCCTTTTAAT | TAATTATCGC  | TGCTAAATGT  | CTTCGTTCCA | ATAAGGTAGT  |
| 5551 | CATATATTGA | TTTATGTACT  | GTTTCAATTA  | AAAAAGGTAA | TTCAAATGAA  |
|      | GTATATAACT | AAATACATGA  | CAAAGTTAAT  | TTTTTCCATT | AAGTTTACTT  |
| 5601 | ATTGTTAAAT | GTAATTAATT  | TTGTTTTCTT  | GATGTTTGTT | TCATCATCTT  |
|      | TAACAATTTA | CATTAATTAA  | AACAAAAGAA  | CTACAAACAA | AGTAGTAGAA  |
| 5651 | CTTTTGCTCA | AGTAATTGAA  | ATGAATAATT  | CGCCTCTGCG | CGATTTCTGT  |
|      | GAAAACGAGT | TCATTAACCT  | TACTTATTAA  | GCGGAGACGC | GCTAAAGCAC  |
| 5701 | ACTTGGTATT | CAAAGCAAAC  | AGGTGAATCT  | GTTATTGTCT | CACCTGATGT  |
|      | TGAACCATAA | GTTTCGTTTG  | TCCACTTAGA  | CAATAACAGA | GTGGACTACA  |

11/39

5751 TAAAGGTACA GTGACTGTAT ATTCTCTGA CGTTAAGCCT GAAAATTTAC  
 ATTTCCATGT CACTGACATA TAAGGAGACT GCAATTCGGA CTTTAAATG

5801 GCAATTTCTT TATCTCTGTT TTACGTGCTA ATAATTTTGA TATGGTTGGC  
 CGTTAAAGAA ATAGAGACAA AATGCACGAT TATTAAACT ATACCAACCG

5851 TCAATTCCTT CCATAATTCA GAAATATAAC CCAAATAGTC AGGATTATAT  
 AGTTAAGGAA GGTATTAAGT CTTTATATTG GGTTTATCAG TCCTAATATA

5901 TGATGAATTG CCATCATCTG ATATTCAGGA ATATGATGAT AATTCCGCTC  
 ACTACTTAAC GGTAGTAGAC TATAAGTCCT TATACTACTA TTAAGGCGAG

5951 CTTCTGGTGG TTTCTTTGTT CCGCAAAATG ATAATGTTAC TCAAACATTT  
 GAAGACCACC AAAGAAACAA GGCGTTTTAC TATTACAATG AGTTTGTAAA

6001 AAAATTAATA ACGTTCGCGC AAAGGATTTA ATAAGGGTTG TAGAATTGTT  
 TTTTAATTAT TGCAAGCGCG TTTCTTAAAT TATTCCTAAC ATCTTAACAA

6051 TGTTAAATCT AATACATCTA AATCCTCAAA TGTATTATCT GTTGATGGTT  
 ACAATTTAGA TTATGTAGAT TTAGGAGTTT ACATAATAGA CAACTACCAA

6101 CTAACCTTATT AGTAGTTAGC GCCCTAAAG ATATTTTAGA TAACCTTCCG  
 GATTGAATAA TCATCAATCG CGGGGATTTT TATAAAATCT ATTGGAAGGC

6151 CAATTTCTTT CTACTGTTGA TTTGCCAACT GACCAGATAT TGATTGAAGG  
 GTTAAAGAAA GATGACAACT AAACGGTTGA CTGGTCTATA ACTAACTTCC

6201 ATTAATTTTC GAGGTTTCTG AAGGTGATGC TTTAGATTTT TCCTTTGCTG  
 TAATTAAAAG CTCCAAGTCG TTCCACTACG AAATCTAAAA AGGAAACGAC

6251 CTGGCTCTCA GCGCGGCACT GTTGCTGGTG GTGTTAATAC TGACCGTCTA  
 GACCGAGAGT CGCGCCGTGA CAACGACCAC CACAATTATG ACTGGCAGAT

6301 ACCTCTGTTT TATCTTCTGC GGGTGGTTCTG TTCGGTATTT TTAACGGCGA  
 TGGAGACAAA ATAGAAGACG CCCACCAAGC AAGCCATAAA AATTGCCGCT

6351 TGTTTTAGGG CTATCAGTTC GCGCATTAAA GACTAATAGC CATTCAAAAA  
 ACAAATCCC GATAGTCAAG CGCGTAATTT CTGATTATCG GTAAGTTTTT

6401 TATTGTCTGT GCCTCGTATT CTTACGCTTT CAGGTCAGAA GGGTTCTATT  
 ATAACAGACA CGGAGCATAA GAATGCGAAA GTCCAGTCTT CCAAGATAA

6451 TCTGTTGGCC AGAATGTCCC TTTTATTACT GGTCTGTGTA CTGGTGAATC  
 AGACAACCGG TCTTACAGGG AAAATAATGA CCAGCACATT GACCACTTAG

6501 TGCCAATGTA AATAATCCAT TTCAGACGGT TGAGCGTCAA AATGTTGGTA  
 ACGGTTACAT TTATTAGGTA AAGTCTGCCA ACTCGCAGTT TTACAACCAT

6551 TTTCTATGAG TGTTTTTCCC GTTGCAATGG CTGGCGGTAA TATTGTTTTA  
 AAAGATACTC ACAAAAAGGG CAACGTTACC GACCGCCATT ATAACAAAAT

12/39

6601 GATATAACCA GTAAGGCCGA TAGTTTGAGT TCTTCTACTC AGGCAAGTGA  
CTATATTGGT CATTCCGGCT ATCAAACCTCA AGAAGATGAG TCCGTTCACT

6651 TGTATTACT AATCAAAGAA GTATTGCGAC AACGGTTAAT TTGCGTGATG  
ACAATAATGA TTAGTTTCTT CATAACGCTG TTGCCAATTA AACGCACCTAC

6701 GTCAGACTCT TTTGCTCGGT GGCCTCACTG ATTACAAAAA CACTTCTCAA  
CAGTCTGAGA AAACGAGCCA CCGGAGTGAC TAATGTTTTT GTGAAGAGTT

6751 GATTCTGGTG TGCCGTTTCT GTCTAAAATC CCTTTAATCG GCCTCCTGTT  
CTAAGACCAC ACGGCAAGGA CAGATTTTAG GGAAATTAGC CGGAGGACAA

6801 TAGCTCCCGT TCTGATTCTA ACGAGGAAAAG CACGTTGTAC GTGCTCGTCA  
ATCGAGGGCA AGACTAAGAT TGCTCCTTTC GTGCAACATG CACGAGCAGT

6851 AAGCAACCAT AGTACGCGCC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT  
TTCGTTGGTA TCATGCGCGG GACATCGCCG CGTAATTCGC GCCGCCACA

6901 GGTGGTTACG CGCAGCGTGA CCGCTACACT TGCCAGCGCC CTAGCGCCCG  
CCACCAATGC GCGTCGCACT GGCGATGTGA ACGGTCGCGG GATCGCGGGC

6951 CTCCTTTTCG TTTCTTCCCT TCCTTTCTCG CCACGTTCTC CGGCTTTCCC  
GAGGAAAGCG AAAGAAGGGA AGGAAAGAGC GGTGCAAGAG GCCGAAAGGG

BamHI

~~~~~

7001 CGTCAAGCTC TAAATCGGGG GATCCCTTTA GGGTTCCGAT TTAGTGCTTT
GCAGTTCGAG ATTTAGCCCC CTAGGGAAAT CCCAAGGCTA AATCACGAAA

7051 ACGGCACCTC GACCTCCAAA AACTTGATTT GGGTGATGGT TCACGTAGTG
TGCCGTGGAG CTGGAGGTTT TTGAACTAAA CCCACTACCA AGTGCATCAC

7101 GGCCATCGCC CTGATAGACG GTTTTTTCGCC CTTTGACGTT GGAGTCCACG
CCGGTAGCGG GACTATCTGC CAAAAGCGG GAAACTGCAA CCTCAGGTGC

7151 TTCTTTAATA GTGGACTCTT GTTCCAAACT GGAACAACAC TCACAACTAA
AAGAAATTAT CACCTGAGAA CAAGGTTTGA CCTTGTTGTG AGTGTTGATT

7201 CTCGGCCTAT TCTTTTGATT TATAAGGATT TTTGTCATTT TCTGCTTACT
GAGCCGGATA AGAAAACATA ATATTCCTAA AAACAGTAAA AGACGAATGA

7251 GGTTAACAAA TAAGCTGATT TAACAAATAT TTAACGCGAA ATTTAACAAA
CCAATTTTTT ATTCGACTAA ATTGTTTATA AATTGCGCTT TAAATTGTTT

7301 ACATTAACGT TTACAATTTA AATATTTGCT TATACAATCA TCCTGTTTTT
TGTAATTGCA AATGTTAAAT TTATAACGA ATATGTTAGT AGGACAAAAA

7351 GGGGCTTTTC TGATTATCAA CCGGGGTACA TATGATTGAC ATGCTAGTTT
CCCCGAAAAG ACTAATAGTT GGCCCCATGT ATACTAAGT TACGATCAAA

13/39

ClaI

~~~~~

7401 TACGATTACC GTTCATCGAT TCTCTTGTTT GCTCCAGACT TTCAGGTAAT  
ATGCTAATGG CAAGTAGCTA AGAGAACAAA CGAGGTCTGA AAGTCCATTA

7451 GACCTGATAG CCTTTGTAGA CCTCTCAAAA ATAGCTACCC TCTCCGGCAT  
CTGGACTATC GGAAACATCT GGAGAGTTTT TATCGATGGG AGAGGCCGTA

7501 GAATTTATCA GCTAGAACGG TTGAATATCA TATTGACGGT GATTTGACTG  
CTTAAATAGT CGATCTTGCC AACTTATAGT ATAAGTCCCA CTAAACTGAC

7551 TCTCCGGCCT TTCTCACCCG TTTGAATCTT TGCCTACTCA TTACTCCGGC  
AGAGGCCGGA AAGAGTGGGC AAAGTTAGAA ACGGATGAGT AATGAGGCCG

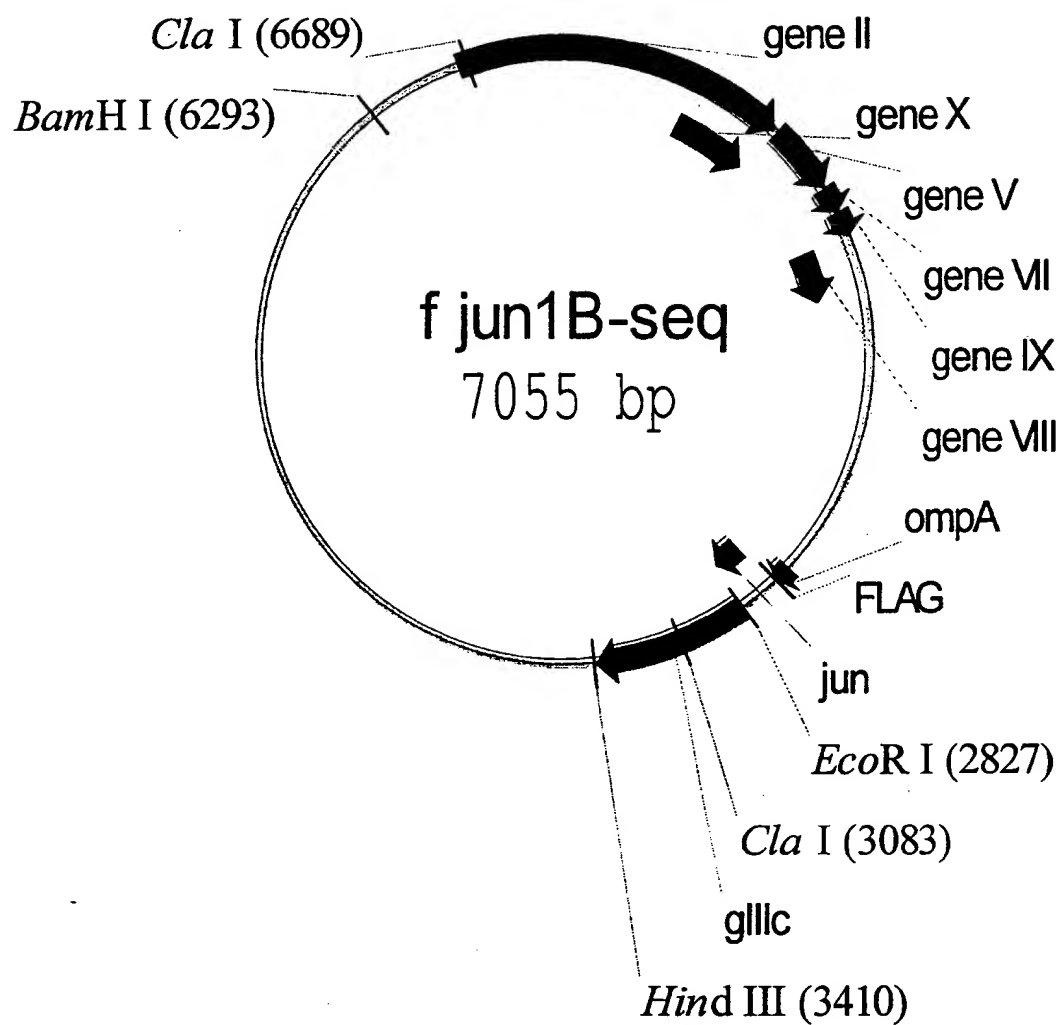
7601 ATTGCATTTA AAATATATGA GGGTTCTAAA AATTTTTATC CCTGCGTTGA  
TAACGTAAAT TTTATATACT CCCAAGATTT TTAAAAATAG GGACGCAACT

7651 AATTAAGGCT TCACCAGCAA AAGTATTACA GGGTCATAAT GTTTTTGGTA  
TTAATTCCGA AGTGGTCGTT TTCATAATGT CCCAGTATTA CAAAAACCAT

7701 CAACCGATTT AGCTTTATGC TCTGAGGCTT TATTGCTTAA TTTTGCTAAC  
GTTGGCTAAA TCGAAATACG AGACTCCGAA ATAACGAATT AAAACGATTG

7751 TCTCTGCCTT GCTTGTACGA TTTATTGGAT GTT  
AGAGACGGAA CGAACATGCT AAATAACCTA CAA

14/39

**Figure 3**

15/39

```

1  AACGCTACTA CCATTAGTAG AATTGATGCC ACCTTTTCAG CTCGCGCCCC
   TTGCGATGAT GGTAATCATC TTAACTACGG TGGAAAAGTC GAGCGCGGGG

51  AAATGAAAAT ATAGCTAAAC AGGTTATTGA CCATTTCGCA AATGTATCTA
   TTTACTTTTA TATCGATTG TCCAATAACT GGTAACGCT TTACATAGAT

101 ATGGTCAAAC TAAATCTACT CGTTCGCAGA ATTGGGAATC AACTGTTACA
   TACCAGTTTG ATTTAGATGA GCAAGCGTCT TAACCCTTAG TTGACAAATG

151 TGGAATGAAA CTTCCAGACA CCGTACTTTA GTTGCAATATT TAAAACATGT
   ACCTTACTTT GAAGGTCTGT GGCATGAAAT CAACGTATAA ATTTTGTACA

201 TGAACCTACAG CACCAGATTC AGCAATTAAG CTCTAAGCCA TCCGCAAAAA
   ACTTGATGTC GTGGTCTAAG TCGTTAATTC GAGATTCGGT AGGCGTTTTT

251 TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTGTCTAA TCCTGACCTG
   ACTGGAGAAT AGTTTTCTCT GTTAATTTCC ATGACAGATT AGGACTGGAC

301 TTGGAATTTG CTTCCGGTCT GGTTTCGCTTT GAGGCTCGAA TTGAAACGCG
   AACCTTAAAC GAAGGCCAGA CCAAGCGAAA CTCCGAGCTT AACTTTGCGC

351 ATATTTGAAG TCTTTCGGGC TTCCTCTTAA TCTTTTTGAT GCAATTCGCT
   TATAAACTTC AGAAAGCCCG AAGGAGAATT AGAAAACTA CGTTAAGCGA

401 TTGCTTCTGA CTATAATAGA CAGGGTAAAG ACCTGATTTT TGATTTATGG
   AACGAAGACT GATATTATCT GTCCCATTTT TGGACTAAAA ACTAAATACC

451 TCATTCTCGT TTTCTGAACT GTTTAAAGCA TTTGAGGGGG ATTCAATGAA
   AGTAAGAGCA AAAGACTTGA CAAATTTTCGTA AAACCCCC TAAGTTACTT

501 TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT AAACATTTTA
   ATAAATACTG CTAAGGCGTC ATAACCTGCG ATAGGTCAGA TTTGTAAAAA

551 CAATTACCCC CTCTGGCAAA ACTTCCTTTG CAAAAGCCTC TCGCTATTTT
   GTTAATGGGG GAGACCGTTT TGAAGGAAAC GTTTTCGGAG AGCGATAAAA

601 GGTTTCTATC GTCGTCTGGT TAATGAGGGT TATGATAGTG TTGCTCTTAC
   CCAAAGATAG CAGCAGACCA ATTACTCCCA ATACTATCAC AACGAGAATG

651 CATGCCTCGT AATTCCTTTT GCGGTTATGT ATCTGCATTA GTTGAGTGTG
   GTACGGAGCA TTAAGGAAAA CCGCAATACA TAGACGTAAT CAACTCACAC

701 GTATTCCTAA ATCTCAATTG ATGAATCTTT CCACCTGTAA TAATGTTGTT
   CATAAGGATT TAGAGTTAAC TACTTAGAAA GGTGGACATT ATTACAACAA

751 CCGTTAGTTC GTTTTATTAA CGTAGATTTT TCCTCCCAAC GTCCTGACTG
   GGCAATCAAG CAAAATAATT GCATCTAAAA AGGAGGGTTG CAGGACTGAC

801 GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA AAATGATTAA
   CATATTACTC GGTCAAGAAT TTTAGCGTAT TCCATTAAGT TTTACTAATT

```



16/39

|      |            |             |             |             |            |
|------|------------|-------------|-------------|-------------|------------|
| 851  | AGTTGAAATT | AAACCGTCTC  | AAGCGCAATT  | TACTACCCGT  | TCTGGTGTTT |
|      | TCAACTTTAA | TTTGGCAGAG  | TTCGCGTTAA  | ATGATGGGCA  | AGACCACAAA |
| 901  | CTCGTCAGGG | CAAGCCTTAT  | TCACTGAATG  | AGCAGCTTTG  | TTACGTTGAT |
|      | GAGCAGTCCC | GTTCGGAATA  | AGTGACTTAC  | TCGTCGAAAC  | AATGCAACTA |
| 951  | TTGGGTAATG | AATATCCGGT  | GCTTGTC AAG | ATTACTCTCG  | ACGAAGGTCA |
|      | AACCCATTAC | TTATAGGCCA  | CGAACAGTTC  | TAATGAGAGC  | TGCTTCCAGT |
| 1001 | GCCAGCGTAT | GCGCCTGGTC  | TGTACACCGT  | GCATCTGTCC  | TCGTTCAAAG |
|      | CGGTTCGATA | CGCGGACCAG  | ACATGTGGCA  | CGTAGACAGG  | AGCAAGTTTC |
| 1051 | TTGGTCAGTT | CGGTTCTCTT  | ATGATTGACC  | GTCTGCGCCT  | CGTTCCGGCT |
|      | AACCAGTCAA | GCCAAGAGAA  | TACTAACTGG  | CAGACGCGGA  | GCAAGGCCGA |
| 1101 | AAGTAACATG | GAGCAGGTCTG | CGGATTTCGA  | CACAATTTAT  | CAGGCGATGA |
|      | TTCATTGTAC | CTCGTCCAGC  | GCCTAAAGCT  | GTGTTAAATA  | GTCCGCTACT |
| 1151 | TACAAATCTC | CGTTGTACTT  | TGTTTCGCGC  | TTGGTATAAT  | CGCTGGGGGT |
|      | ATGTTTAGAG | GCAACATGAA  | ACAAAGCGCG  | AACCATATTA  | GCGACCCCCA |
| 1201 | CAAAGATGAG | TGTTTTAGTG  | TATTCTTTTCG | CCTCTTTTCGT | TTTAGGTTGG |
|      | GTTTCTACTC | ACAAAATCAC  | ATAAGAAAGC  | GGAGAAAGCA  | AAATCCAACC |
| 1251 | TGCCTTCGTA | GTGGCATTAC  | GTATTTTACC  | CGTTTAATGG  | AAACTTCCTC |
|      | ACGGAAGCAT | CACCGTAATG  | CATAAAATGG  | GCAAATTACC  | TTTGAAGGAG |
| 1301 | ATGCGTAAGT | CTTTAGTCCT  | CAAAGCCTCC  | GTAGCCGTTG  | CTACCCTCGT |
|      | TACGCATTCA | GAAATCAGGA  | GTTTCGGAGG  | CATCGGCAAC  | GATGGGAGCA |
| 1351 | TCCGATGCTG | TCTTTCGCTG  | CTGAGGGTGA  | CGATCCCGCA  | AAAGCGGCCT |
|      | AGGCTACGAC | AGAAAGCGAC  | GACTCCCACT  | GCTAGGGCGT  | TTTCGCCGGA |
| 1401 | TTGACTCCCT | GCAAGCCTCA  | GCGACCGAAT  | ATATCGGTTA  | TGCGTG GCG |
|      | AACTGAGGGA | CGTTCGGAGT  | CGCTGGCTTA  | TATAGCCAAT  | ACGCACCCGC |
| 1451 | ATGGTTGTTG | TCATTGTCGG  | CGCAACTATC  | GGTATCAAGC  | TGTTTAAGAA |
|      | TACCAACAAC | AGTAACAGCC  | GCGTTGATAG  | CCATAGTTTCG | ACAAATTCTT |
| 1501 | ATTCACCTCG | AAAGCAAGCT  | GATAAAGGAG  | GTTTCTCGAT  | CGAGACGTTN |
|      | TAAGTGGAGC | TTTCGTTCTGA | CTATTTCTCTC | CAAAGAGCTA  | GCTCTGCAAN |
| 1551 | NNNGAGGTTC | CAACTTTTAC  | CATAATGAAA  | TAAGATCACT  | ACCGGGCGTA |
|      | NNNCTCCAAG | GTTGAAAGTG  | GTATTACTTT  | ATTCTAGTGA  | TGGCCCGCAT |
| 1601 | TTTTTTGAGT | TATCGAGATT  | TTCAGGAGCT  | AAGGAAGCTA  | AAATGGAGAA |
|      | AAAAAATCA  | ATAGCTCTAA  | AAGTCCTCGA  | TTCCTTCGAT  | TTTACCTCTT |
| 1651 | AAAAATCACT | GGATATACCA  | CCGTTGATAT  | ATCCCAATGG  | CATCGTAAAG |
|      | TTTTTAGTGA | CCTATATGGT  | GGCAACTATA  | TAGGGTTACC  | GTAGCATTTT |

17/39

|      |             |             |            |            |            |
|------|-------------|-------------|------------|------------|------------|
| 1701 | AACATTTTGA  | GGCATTTCAG  | TCAGTTGCTC | AATGTACCTA | TAACCAGACC |
|      | TTGTAAAACT  | CCGTAAAGTC  | AGTCAACGAG | TTACATGGAT | ATTGGTCTGG |
| 1751 | G TTCAGCTGG | ATATTACGGC  | CTTTTTAAAG | ACCGTAAAGA | AAAATAAGCA |
|      | CAAGTCGACC  | TATAATGCCG  | GAAAAATTTT | TGGCATTCT  | TTTTATTCTG |
| 1801 | CAAGTTTTAT  | CCGGCCTTTA  | TTCACATTCT | TGCCCCCCTG | ATGAATGCTC |
|      | GTTCAAAATA  | GGCCGGAAT   | AAGTGTAAGA | ACGGGCGGAC | TACTTACGAG |
| 1851 | ATCCGGAGTT  | CCGTATGGCA  | ATGAAAGACG | GTGAGCTGGT | GATATGGGAT |
|      | TAGGCCTCAA  | GGCATAACCGT | TACTTTCTGC | CACTCGACCA | CTATACCCTA |
| 1901 | AGTGTTCAAC  | CTTGTTACAC  | CGTTTTCCAT | GAGCAAACTG | AAACGTTTTT |
|      | TCACAAGTGG  | GAACAATGTG  | GCAAAAGGTA | CTCGTTTGAC | TTTGCAAAAG |
| 1951 | ATCGCTCTGG  | AGTGAATACC  | ACGACGATTT | CCGGCAGTTT | CTACACATAT |
|      | TAGCGAGACC  | TCACTTATGG  | TGCTGCTAAA | GGCCGTCAA  | GATGTGTATA |
| 2001 | ATTCGCAAGA  | TGTGGCGTGT  | TACGGTGAAA | ACCTGGCCTA | TTTCCCTAAA |
|      | TAAGCGTTCT  | ACACCGCACA  | ATGCCACTTT | TGGACCGGAT | AAAGGGATTT |
| 2051 | GGGTTTATTG  | AGAATATGTT  | TTTCGTCTCA | GCCAATCCCT | GGGTGAGTTT |
|      | CCCAAATAAC  | TCTTATACAA  | AAAGCAGAGT | CGGTTAGGGA | CCCACTCAAA |
| 2101 | CACCAGTTTT  | GATTTAAACG  | TAGCCAATAT | GGACAACTTC | TTCGCCCCCG |
|      | GTGGTCAAAA  | CTAAATTTGC  | ATCGGTTATA | CCTGTTGAAG | AAGCGGGGGC |
| 2151 | TTTTCACTAT  | GGGCAAATAT  | TATACGCAAG | GCGACAAGGT | GCTGATGCCG |
|      | AAAAGTGATA  | CCCGTTTATA  | ATATGCGTTC | CGCTGTTCCA | CGACTACGGC |
| 2201 | CTGGCGATT   | AGGTTTATCA  | TGCCGTTTGT | GATGGCTTCC | ATGTCGGCAG |
|      | GACCGCTAAG  | TCCAAGTAGT  | ACGGCAAACA | CTACCGAAGG | TACAGCCGTC |
| 2251 | AATGCTTAAT  | GAATTACAAC  | AGTACTGCGA | TGAGTGGCAG | GGCGGGGCGT |
|      | TTACGAATTA  | CTTAATGTTG  | TCATGACGCT | ACTCACCCTC | CCGCCCCGCA |
| 2301 | AATTTTTTTA  | AGGCAGTTAT  | TGGTGCCCTT | AAACGCCTGG | TGCTAGCCTG |
|      | TTAAAAAAAT  | TCCGTCAATA  | ACCACGGGAA | TTTGCGGACC | ACGATCGGAC |
| 2351 | AGGCCAGTTT  | GCTCAGGCTC  | TCCCCGTGGA | GGTAATAATT | GCTCGACCGA |
|      | TCCGGTCAAA  | CGAGTCCGAG  | AGGGGCACCT | CCATTATTAA | CGAGCTGGCT |
| 2401 | TAAAAGCGGC  | TTCCTGACAG  | GAGGCCGTTT | TGTTTTGCAG | CCCACCTCAA |
|      | ATTTTCGCCG  | AAGGACTGTC  | CTCCGGCAAA | ACAAAACGTC | GGGTGGAGTT |
| 2451 | CGCAATTAAT  | GTGAGTTAGC  | TCACTCATT  | GGCACCCAG  | GCTTTACACT |
|      | GCGTTAATTA  | CACTCAATCG  | AGTGAGTAAT | CCGTGGGGTC | CGAAATGTGA |
| 2501 | TTATGCTTCC  | GGCTCGTATG  | TTGTGTGGAA | TTGTGAGCGG | ATAACAATTT |
|      | AATACGAAGG  | CCGAGCATAC  | AACACACCTT | AACACTCGCC | TATTGTAAAA |

18/39

2551 CACACAGGAA ACAGCTATGA CCATGATTAC GAATTTCTAG ATAACGAGGG  
GTGTGTCCTT TGTCGATACT GGTACTAATG CTTAAAGATC TATTGCTCCC

2601 CAAAAAATGA AAAAGACAGC TATCGCGATT GCAGTGGCAC TGGCTGGTTT  
GTTTTTTACT TTTTCTGTCT ATAGCGCTAA CGTCACCGTG ACCGACCAAA

2651 CGCTACCGTA GCGCAGGCCG ACTACAAAGA TGTCGACGCC GGTGGTCGGA  
GCGATGGCAT CGCGTCCGGC TGATGTTTCT ACAGCTGCGG CCACCAGCCT

2701 TCGCCCGGCT AGAGGAAAAA GTGAAAACCT TGAAAGCGCA AACTCCGAG  
AGCGGGCCGA TCTCCTTTTT CACTTTTGGA ACTTTCGCGT TTTGAGGCTC

2751 CTGGCGTCCA CGGCCAATCAT GCTCAGGGAA CAGGTGGCAC AGCTTAAACA  
GACCGCAGGT GCCGGTTGTA CGAGTCCCTT GTCCACCGTG TCGAATTTGT

EcoRI

~~~~~

2801 GAAAGTCATG AACCACGGTG GTGCCGAATT CAATGCTGGC GGCGGCTCTG
CTTTCAGTAC TTGGTGCCAC CACGGCTTAA GTTACGACCG CCGCCGAGAC

2851 GTGGTGGTTC TGGTGGCGGC TCTGAGGGTG GTGGCTCTGA GGGTGGCGGT
CACCACCAAG ACCACCGCCG AGACTCCAC CACCGAGACT CCCACGCCA

2901 TCTGAGGGTG GCGGCTCTGA GGGAGGCGGT TCCGGTGGTG GCTCTGGTTC
AGACTCCAC CGCCGAGACT CCTCCGCCA AGGCCACCAC CGAGACCAAG

2951 CGGTGATTTT GATTATGAAA AGATGGCAA CGCTAATAAG GGGGCTATGA
GCCACTAAAA CTAATACTTT TCTACCGTTT GCGATTATTC CCCCATACT

3001 CCGAAAAATGC CGATGAAAAC GCGCTACAGT CTGACGCTAA AGGCAAACCT
GGCTTTTACG GCTACTTTTG CGCGATGTCA GACTGCGATT TCCGTTTGAA

ClaI

~~~~~

3051 GATTCTGTCTG CTAAGACAGC GATGACTAAT GCCACGACGA TAGCTACCAA AGTAACCACT

3101 CGTTTCCGGC CTTGCTAATG GTAATGGTGC TACTGGTGAT TTTGCTGGCT  
GCAAAGGCCG GAACGATTAC CATTACCACG ATGACCACTA AAACGACCGA

3151 CTAATTCCCA AATGGCTCAA GTCGGTGACG GTGATAATTC ACCTTTAATG  
GATTAAGGGT TTACCGAGTT CAGCCACTGC CACTATTAAG TGGAAATTAC

3201 AATAATTTCC GTCAATATTT ACCTTCCCTC CCTCAATCGG TTGAATGTCTG  
TTATTAAAGG CAGTTATAAA TGGAAAGGAG GGAGTTAGCC AACTTACAGC

3251 CCCTTTTGTC TTTAGCGCTG GTAAACCATA TGAATTTTCT ATTGATTGTG  
GGGAAAACAG AAATCGCGAC CATTTGGTAT ACTTAAAGA TAACTAACAC

3301 ACAAATAAAA CTTATTCCGT GGTGTCTTTG CGTTTCTTTT ATATGTTGCC  
TGTTTTATTT GAATAAGGCA CCACAGAAAC GCAAAGAAAA TATACAACGG

19/39

3351 ACCTTTATGT ATGTATTTTC TACGTTTGCT AACATACTGC GTAATAAGGA  
TGAAAATACA TACATAAAAG ATGCAAACGA TTGTATGACG CATTATTCCT

HindIII

~~~~~

3401 GTCTTGATAA GCTTCGAGAA ATTCACCTCG AAAGCAAGCT GATAAACCGA
CAGAACTATT CGAAGCTCTT TAAGTGGAGC TTTCGTTCTGA CTATTGCGCT

3451 TACAATTAAA GGCTCCTTTT GGAGCCTTTT TTTTGGAGA ATTAATTCAA
ATGTTAATTT CCGAGGAAAA CCTCGGAAAA AAAAACCTCT TAATTAAGTT

3501 TCATGCCAGT TCTTTTGGGT ATTCCGTTAT TATTGCGTTT CCTCGGTTTC
AGTACGGTCA AGAAAACCCA TAAGGCAATA ATAACGCAA GGAGCCAAAG

3551 CTTCTGGTAA CTTTGTTCGG CTATCTGCTT ACTTTCCTTA AAAAGGGCTT
GAAGACCATT GAAACAAGCC GATAGACGAA TGAAAGGAAT TTTTCCCGAA

3601 CGGTAAGATA GCTATTGCTA TTTCATTGTT TCTTGCTCTT ATTATTGGGC
GCCATTCTAT CGATAACGAT AAAGTAACAA AGAACGAGAA TAATAACCCG

3651 TTAACCTCAAT TCTTGTGGGT TATCTCTCTG ATATTAGCGC ACAATTACCC
AATTGAGTTA AGAACACCCA ATAGAGAGAC TATAATCGCG TGTTAATGGG

3701 TCTGATTTTG TTCAGGGCGT TCAGTTAATT CTCCCGTCTA ATGCGCTTCC
AGACTAAAC AAGTCCCGCA AGTCAATTAA GAGGGCAGAT TACGCGAAGG

3751 CTGTTTTTAT GTTATTCTCT CTGTAAAGGC TGCTATTTTC ATTTTGGACG
GACAAAATA CAATAAGAGA GACATTTCCG ACGATAAAAG TAAAACTGC

3801 TTAAACAAAA AATCGTTTCT TATTTGGATT GGGATAAATA AATATGGCTG
AATTGTTTT TTAGCAAAGA ATAAACCTAA CCCTATTTAT TTATACCGAC

3851 TTTATTTTGT AACTGGCAAA TTAGGCTCTG GAAAGACGCT CGTTAGCGTT
AAATAAAACA TTGACCGTTT AATCCGAGAC CTTTCTGCGA GCAATCGCAA

3901 GGTAAGATTC AGGATAAAAT TGTAGCTGGG TGCAAAATAG CAACTAATCT
CCATTCTAAG TCCTATTTTA ACATCGACCC ACGTTTTATC GTTGATTAGA

3951 TGATTTAAGG CTTCAAAACC TCCCGCAAGT CGGGAGGTTC GCTAAAACGC
ACTAAATTCC GAAGTTTTGG AGGGCGTTCA GCCCTCCAAG CGATTTTGCG

4001 CTCGCGTTCT TAGAATACCG GATAAGCCTT CTATTTCTGA TTTGCTTGCT
GAGCGCAAGA ATCTTATGGC CTATTCGGAA GATAAAGACT AAACGAACGA

4051 ATTGGTCGTG GTAATGATTC CTACGACGAA AATAAAAACG GTTTGCTTGT
TAACCAGCAC CATTACTAAG GATGCTGCTT TTATTTTTCG CAAACGAACA

4101 TCTTGATGAA TGCGGTACTT GGTTTAATAC CCGTTCATGG AATGACAAGG
AGAACTACTT ACGCCATGAA CCAAATTATG GGCAAGTACC TTACTGTTCC

20/39

4151 AAAGACAGCC GATTATTGAT TGGTTTCTTC ATGCTCGTAA ATTGGGATGG
 TTTCTGTCGG CTAATAACTA ACCAAAGAAG TACGAGCATT TAACCCTACC
 4201 GATATTATTT TTCTTGTTCA GGATTTATCT ATTGTTGATA AACAGGCGCG
 CTATAATAAA AAGAACAAGT CCTAAATAGA TAACAACTAT TTGTCCGCGC
 4251 TTCTGCATTA GCTGAACACG TTGTTTATTG TCGCCGTCTG GACAGAATTA
 AAGACGTAAT CGACTTGTGC AACAAATAAC AGCGGCAGAC CTGTCTTAAT
 4301 CTTTACCCTT TGTCGGCACT TTATATTCTC TTGTTACTGG CTCAAAAATG
 GAAATGGGAA ACAGCCGTGA AATATAAGAG AACAATGACC GAGTTTTTAC
 4351 CCTCTGCCTA AATTACATGT TGGTGTGTGT AAATATGGTG ATTCTCAATT
 GGAGACGGAT TTAATGTACA ACCACAACAA TTTATACCAC TAAGAGTTAA
 4401 AAGCCCTACT GTTGAGCGTT GGCTTTTATAC TGGTAAGAAT TTATATAACG
 TTCGGGATGA CAACTCGCAA CCGAAATATG ACCATTCTTA AATATATTGC
 4451 CATATGACAC TAAACAGGCT TTTTCCAGTA ATTATGATTC AGGTGTTTAT
 GTATACTGTG ATTTGTCCGA AAAAGGTCAT TAATACTAAG TCCACAAATA
 4501 TCATATTTAA CCCCTTATTT ATCACACGGT CGGTATTTCA AACCATTAAA
 AGTATAAATT GGGGAATAAA TAGTGTGCCA GCCATAAAGT TTGGTAATTT
 4551 TTTAGGTCAG AAGATGAAAT TAACTAAAT ATATTTGAAA AAGTTTTCTC
 AAATCCAGTC TTCTACTTTA ATTGATTTTA TATAAACTTT TTCAAAAGAG
 4601 GCGTTCTTTG TCTTGCGATA GGATTTGCAT CAGCATTTAC ATATAGTTAT
 CGCAAGAAAC AGAACGCTAT CCTAAACGTA GTCGTAAATG TATATCAATA
 4651 ATAACCCAAC CTAAGCCGGA GGTAAAAAAG GTAGTCTCTC AGACCTATGA
 TATTGGGTTG GATTGGGCTT CCAATTTTTT CATCAGAGAG TCTGGATACT
 4701 TTTTGATAAA TTTACTATTG ACTCTTCTCA GCGTCTTAAT CTAAGCTATC
 AAAACTATTT AAGTGATAAC TGAGAAGAGT CGCAGAATTA GATTGATAG
 4751 GCTATGTTTT CAAGGATTCT AAGGGAAAAT TAATTAATAG CGACGATTTA
 CGATACAAAA GTTCCTAAGA TTCCCTTTTA ATTAATTATC GCTGCTAAAT
 4801 CAGAAGCAAG GTTATTCCAT CACATATATT GATTTATGTA CTGTTTCAAT
 GTCTTCGTTC CAATAAGGTA GTGTATATAA CTAAATACAT GACAAAGTTA
 4851 TAAAAAAGGT AATTCAAATG AAATTGTTAA ATGTAATTAA TTTTGTTTTT
 ATTTTTTCCA TTAAGTTTAC TTTAACAATT TACATTAATT AAAACAAAAG
 4901 TTGATGTTTG TTTCATCATC TTCTTTTGCT CAAGTAATTG AAATGAATAA
 AACTACAAAC AAAGTAGTAG AAGAAAACGA GTTCATTAAC TTTACTTATT
 4951 TTCGCCTCTG CGCGATTTCG TGACTTGGTA TTCAAAGCAA ACAGGTGAAT
 AAGCGGAGAC GCGCTAAAGC ACTGAACCAT AAGTTTCGTT TGTCCACTTA

21/39

5001 CTGTTATTGT CTCACCTGAT GTTAAAGGTA CAGTGACTGT ATATTCCTCT
 GACAATAACA GAGTGGACTA CAATTTCCAT GTCACTGACA TATAAGGAGA
 5051 GACGTTAAGC CTGAAAATTT ACGCAATTTT TTTATCTCTG TTTTACGTGC
 CTGCAATTCG GACTTTTAAA TGC GTTAAAG AAATAGAGAC AAAATGCACG
 5101 TAATAATTTT GATATGGTTG GCTCAATTCC TTCCATAATT CAGAAATATA
 ATTATTAAAA CTATACCAAC CGAGTTAAGG AAGGTATTAA GTCTTTATAT
 5151 ACCCAAATAG TCAGGATTAT ATTGATGAAT TGCCATCATC TGATATTCAG
 TGGGTTTATC AGTCCTAATA TAACTACTTA ACGGTAGTAG ACTATAAGTC
 5201 GAATATGATG ATAATTCGCG TCCTTCTGGT GGTTTCTTTG TTCCGCAAAA
 CTTATACTAC TATTAAGGCG AGGAAGACCA CCAAAGAAAC AAGGCGTTTT
 5251 TGATAATGTT ACTCAAACAT TAAAATTAA TAACGTTCGC GCAAAGGATT
 ACTATTACAA TGAGTTTGTA AATTTTAATT ATTGCAAGCG CGTTTCCTAA
 5301 TAATAAGGGT TGTAGAATTG TTTGTAAAT CTAATACATC TAAATCCTCA
 ATTATTCCTA ACATCTTAAC AAACAATTTA GATTATGTAG ATTTAGGAGT
 5351 AATGTATTAT CTGTTGATGG TTCTAACTTA TTAGTAGTTA GCGCCCCTAA
 TTACATAATA GACAACTACC AAGATTGAAT AATCATCAAT CGCGGGGATT
 5401 AGATATTTTA GATAACCTTC CGCAATTTCT TTCTACTGTT GATTTGCCAA
 TCTATAAAAT CTATTGGAAG GCGTTAAAGA AAGATGACAA CTAAACGGTT
 5451 CTGACCAGAT ATTGATTGAA GGATTAATTT TCGAGGTTCA GCAAGGTGAT
 GACTGGTCTA TAACTAACTT CCTAATTAAA AGCTCCAAGT CGTTCCACTA
 5501 GCTTTAGATT TTTCTTTTGC TGCTGGCTCT CAGCGCGGCA CTGTTGCTGG
 CGAAATCTAA AAAGGAAACG ACGACCGAGA GTCGCGCCGT GACAACGACC
 5551 TGGTGTTAAT ACTGACCGTC TAACCTCTGT TTTATCTTCT GCGGGTGGTT
 ACCACAATTA TGA CTGGCAG ATTGGAGACA AAATAGAAGA CGCCACCAA
 5601 CGTTCGGTAT TTTTAACGGC GATGTTTTAG GGCTATCAGT TCGCGCATT
 GCAAGCCATA AAAATTGCCG CTACAAAATC CCGATAGTCA AGCGCGTAAT
 5651 AAGACTAATA GCCATTCAAA AATATTGTCT GTGCCTCGTA TTCTTACGCT
 TTCTGATTAT CGGTAAGTTT TTATAACAGA CACGGAGCAT AAGAATGCGA
 5701 TTCAGGTCAG AAGGGTTCTA TTTCTGTTGG CCAGAAATGTC CCTTTTATTA
 AAGTCCAGTC TTCCCAAGAT AAAGACAACC GGTCTTACAG GGAAATAAT
 5751 CTGGTCGTGT AACTGGTGAA TCTGCCAATG TAAATAATCC ATTTAGACG
 GACCAGCACA TTGACCACTT AGACGGTTAC ATTTATTAGG TAAAGTCTGC
 5801 GTTGAGCGTC AAAATGTTGG TATTTCTATG AGTGTTTTTC CCGTTGCAAT
 CAACTCGCAG TTTTACAACC ATAAAGATAC TCACAAAAAG GGCAACGTTA

22/39

5851 GGCTGGCGGT AATATTGTTT TAGATATAAC CAGTAAGGCC GATAGTTTGA
CCGACCGCCA TTATAACAAA ATCTATATTG GTCATTCCGG CTATCAAAC

5901 GTTCTTCTAC TCAGGCAAGT GATGTTATTA CTAATCAAAG AAGTATTGCG
CAAGAAGATG AGTCCGTTCA CTACAATAAT GATTAGTTTC TTCATAACGC

5951 ACAACGGTTA ATTTGCGTGA TGGTCAGACT CTTTTGCTCG GTGGCCTCAC
TGTTGCCAAT TAAACGCACT ACCAGTCTGA GAAAACGAGC CACCGGAGTG

6001 TGATTACAAA AACACTTCTC AAGATTCTGG TGTGCCGTTT CTGTCTAAAA
ACTAATGTTT TTGTGAAGAG TTCTAAGACC ACACGGCAAG GACAGATTTT

6051 TCCCTTTAAT CGGCCTCCTG TTTAGCTCCC GTTCTGATTC TAACGAGGAA
AGGGAAATTA GCCGGAGGAC AAATCGAGGG CAAGACTAAG ATTGCTCCTT

6101 AGCACGTTGT ACGTGCTCGT CAAAGCAACC ATAGTACGCG CCCTGTAGCG
TCGTGCAACA TGCACGAGCA GTTTCGTTGG TATCATGCGC GGGACATCGC

6151 GCGCATTAAAG CGCGGCGGGT GTGGTGGTTA CGCGCAGCGT GACCGCTACA
CGCGTAATTC GCGCCGCCCA CACCACCAAT GCGCGTCGCA CTGGCGATGT

6201 CTTGCCAGCG CCCTAGCGCC CGCTCCTTTC GCTTTCCTCC CTTCTTTTCT
GAACGGTCGC GGGATCGCGG GCGAGGAAAG CGAAAGAAGG GAAGGAAAGA

BamHI

~~~~~

6251 CGCCACGTTT TCCGGCTTTC CCCGTCAAGC TCTAAATCGG GGGATCCCTT  
GCGGTGCAAG AGGCCGAAAG GGGCAGTTTC AGATTTAGCC CCCTAGGGAA

6301 TAGGGTTCCG ATTTAGTGCT TTACGGCACC TCGACCTCCA AAAACTTGAT  
ATCCAAGGC TAAATCACGA AATGCCGTGG AGCTGGAGGT TTTTGAAC

6351 TTGGGTGATG GTTCACGTAG TGGGCCATCG CCCTGATAGA CGGTTTTTCG  
AACCCACTAC CAAGTGCATC ACCCGGTAGC GGGACTATCT GCCAAAAAGC

6401 CCCTTTGACG TTGGAGTCCA CGTTCTTTAA TAGTGGACTC TTGTTCCAAA  
GGGAAACTGC AACCTCAGGT GCAAGAAATT ATCACCTGAG AACAAAGGTTT

6451 CTGGAACAAC ACTCACAAC AACTCGGCCT ATTCTTTTGA TTTATAAGGA  
GACCTTGTTG TGAGTGTTGA TTGAGCCGGA TAAGAAAAC AAATATTCTT

6501 TTTTGTGTCAT TTTCTGCTTA CTGGTTAAAA AATAAGCTGA TTTAACAAAT  
AAAAACAGTA AAAGACGAAT GACCAATTTT TTATTCGACT AAATTGTTTA

6551 ATTTAACGCG AAATTTAACA AAACATTAAC GTTTACAATT TAAATATTTG  
TAAATTGCGC TTAAATTTGT TTTGTAATTG CAAATGTAA ATTTATAAAC

6601 CTTATACAAT CATCCTGTTT TTGGGGCTTT TCTGATTATC AACCGGGGTA  
GAATATGTTA GTAGGACAAA AACCCGAAA AGACTAATAG TTGGCCCCAT

23/39

Clai

-----

6651 CATATGATTG ACATGCTAGT TTTACGATTA CCGTTCATCG ATTCTCTTGT  
GTATACTAAC TGTACGATCA AAATGCTAAT GGCAAGTAGC TAAGAGAACA

6701 TTGCTCCAGA CTTTCAGGTA ATGACCTGAT AGCCTTTGTA GACCTCTCAA  
AACGAGGTCT GAAAGTCCAT TACTGGACTA TCGGAAACAT CTGGAGAGTT

6751 AAATAGCTAC CCTCTCCGGC ATGAATTTAT CAGCTAGAAC GGTGGAATAT  
TTTATCGATG GGAGAGGCCG TACTTAAATA GTCGATCTTG CCAACTTATA

6801 CATATTGACG GTGATTTGAC TGTCTCCGGC CTTTCTCACC CGTTTGAATC  
GTATAACTGC CACTAAACTG ACAGAGGCCG GAAAGAGTGG GCAAACCTTAG

6851 TTTGCCTACT CATTACTCCG GCATTGCATT TAAAATATAT GAGGGTTCTA  
AAACGGATGA GTAATGAGGC CGTAACGTAA ATTTTATATA CTCCCAAGAT

6901 AAAATTTTTA TCCCTGCGTT GAAATTAAGG CTTCACCAGC AAAAGTATTA  
TTTTAAAAAT AGGGACGCAA CTTTAATTCC GAAGTGGTCG TTTTCATAAT

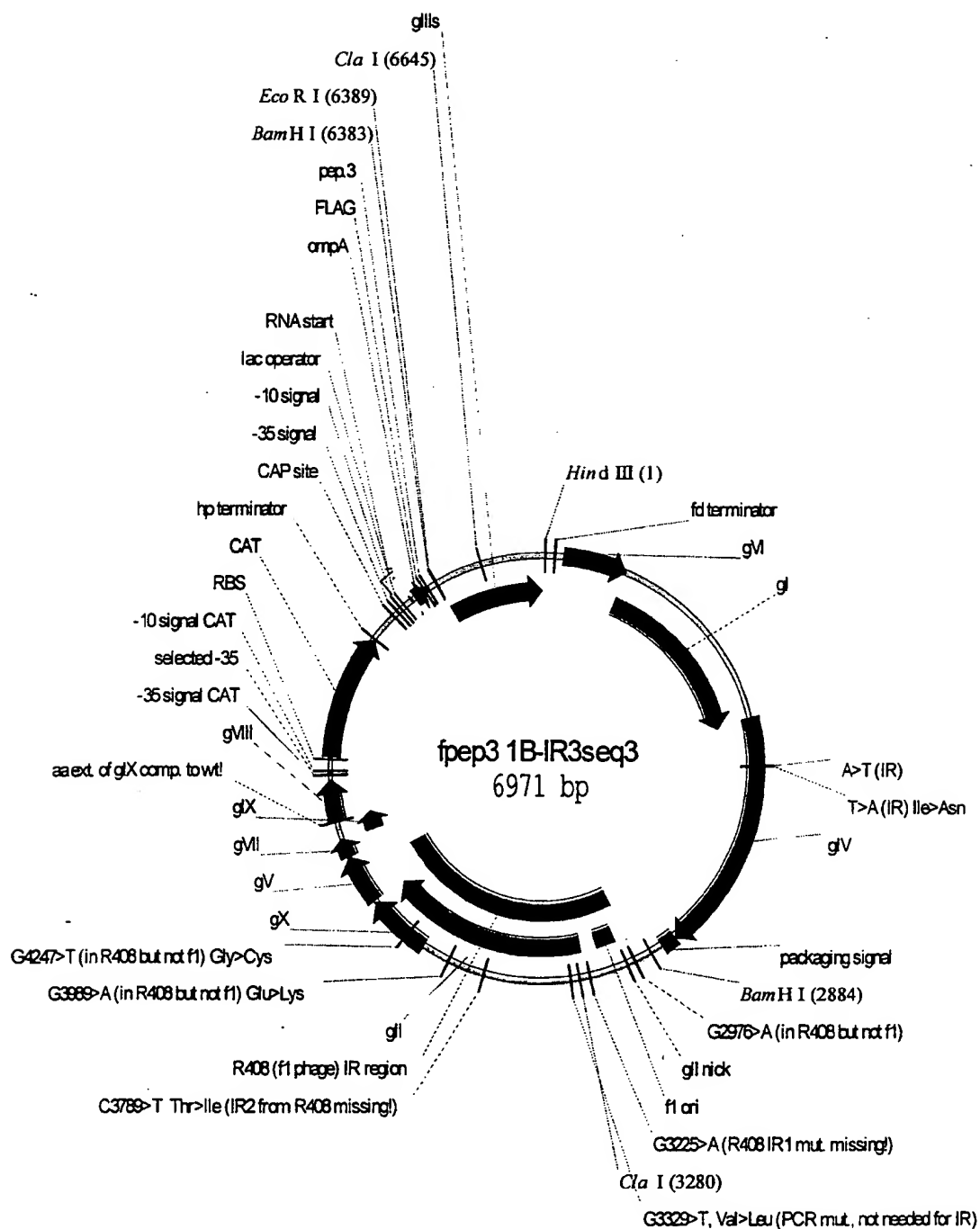
6951 CAGGGTCATA ATGTTTTTGG TACAACCGAT TTAGCTTTAT GCTCTGAGGC  
GTCCCAGTAT TACAAAAACC ATGTTGGCTA AATCGAAATA CGAGACTCCG

7001 TTTATTGCTT AATTTTGCTA ACTCTCTGCC TTGCTTGTAC GATTTATTGG  
AAATAACGAA TTAAACGAT TGAGAGACGG AACGAACATG CTAAATAACC

7051 ATGTT  
TACAA



### Figure 4



25/39

HindIII

~~~~~

1	AGCTTCGAGA	AATTCACCTC	GAAAGCAAGC	TGATAAACCG	ATACAATTAA
	TCGAAGCTCT	TTAAGTGGAG	CTTTCGTTCG	ACTATTTGGC	TATGTTAATT
51	AGGCTCCTTT	TGGAGCCTTT	TTTTTTGGAG	AATTAATTCA	ATCATGCCAG
	TCCGAGGAAA	ACCTCGGAAA	AAAAAACCTC	TTAATTAAGT	TAGTACGGTC
101	TTCTTTTGGG	TATTCGGTTA	TTATTGCGTT	TCCTCGGTTT	CCTTCTGGTA
	AAGAAAACCC	ATAAGGCAAT	AATAACGCAA	AGGAGCCAAA	GGAAGACCAT
151	ACTTTGTTCG	GCTATCTGCT	TACTTTCCTT	AAAAAGGGCT	TCGGTAAGAT
	TGAAACAAGC	CGATAGACGA	ATGAAAGGAA	TTTTTCCCGA	AGCCATTCTA
201	AGCTATTGCT	ATTTCAATTG	TTCTTGCTCT	TATTATTGGG	CTTAACCTCA
	TCGATAACGA	TAAAGTAACA	AAGAACGAGA	ATAATAACCC	GAATTGAGTT
251	TTCTTGTGGG	TTATCTCTCT	GATATTAGCG	CACAATTACC	CTCTGATTTT
	AAGAACACCC	AATAGAGAGA	CTATAATCGC	GTGTTAATGG	GAGACTAAAA
301	G TTCAGGGCG	TT CAGTTAAT	TCTCCCGTCT	AATGCGCTTC	CCTGTTTTTA
	CAAGTCCCGC	AAGTCAATTA	AGAGGGCAGA	TTACGCGAAG	GGACAAAAAT
351	TGTTATTCTC	TCTGTAAAGG	CTGCTATTTT	CATTTTTGAC	GTAAACAAA
	ACAATAAGAG	AGACATTTCC	GACGATAAAA	GTAAAACTG	CAATTTGTTT
401	AAATCGTTTC	TTATTTGGAT	TGGGATAAAAT	AAATATGGCT	GTTTATTTTG
	TTTAGCAAAG	AATAAACCTA	ACCCTATTTA	TTTATACCGA	CAAAATAAAC
451	TAAGTGGCAA	ATTAGGCTCT	GGAAAGACGC	TCGTTAGCGT	TGGTAAGATT
	ATTGACCGTT	TAATCCGAGA	CCTTTCTGCG	AGCAATCGCA	ACCATTTCTA
501	CAGGATAAAA	TTGTAGCTGG	GTGCAAAATA	GCAACTAATC	TTGATTTAAG
	GTCCTATTTT	AACATCGACC	CACGTTTTTAT	CGTTGATTAG	AACTAAATTC
551	GCTTCAAAAC	CTCCCGCAAG	TCGGGAGGTT	CGCTAAAACG	CCTCGCGTTC
	CGAAGTTTTG	GAGGGCGTTC	AGCCCTCCAA	GCGATTTTGC	GGAGCGCAAG
601	TTAGAATACC	GGATAAGCCT	TCTATTTCTG	ATTTGCTTGC	TATTGGTCGT
	AATCTTATGG	CCTATTCGGA	AGATAAAGAC	TAAACGAACG	ATAACCAGCA
651	GGTAATGATT	CCTACGACGA	AAATAAAAAAC	GGTTTGCTTG	TTCTTGATGA
	CCATTACTAA	GGATGCTGCT	TTTATTTTTG	CCAAACGAAC	AAGAACTACT
701	ATGCGGTACT	TGGTTTAATA	CCCGTTCATG	GAATGACAAG	GAAAGACAGC
	TACGCCATGA	ACCAAATTAT	GGGCAAGTAC	CTTACTGTTC	CTTTCTGTGC
751	CGATTATTGA	TTGGTTTCTT	CATGCTCGTA	AATTGGGATG	GGATATTATT
	GCTAATAACT	AACCAAAGAA	GTACGAGCAT	TTAACCTTAC	CCTATAATAA

26/39

801	TTTCTTG TTC	AGGATTTATC	TATTGTTGAT	AAACAGGCGC	GTTCTGCATT
	AAAGAACAAG	TCCTAAATAG	ATAACAATA	TTTGTCCGCG	CAAGACGTAA
851	AGCTGAACAC	GTTGTTTATT	GTCGCCGTCT	GGACAGAATT	ACTTTACCCT
	TCGACTTGTG	CAACAAATAA	CAGCGGCAGA	CCTGTCTTAA	TGAAATGGGA
901	TTGTCGGCAC	TTTATATTCT	CTTGTTACTG	GCTCAAAAAT	GCCTCTGCCT
	AACAGCCGTG	AAATATAAGA	GAACAATGAC	CGAGTTTTTA	CGGAGACGGA
951	AAATTACATG	TTGGTGTGTG	TAAATATGGT	GATTCTCAAT	TAAGCCCTAC
	TTTAATGTAC	AACCACAACA	ATTTATACCA	CTAAGAGTTA	ATTCGGGATG
1001	TGTTGAGCGT	TGGCTTTATA	CTGGTAAGAA	TTTATATAAC	GCATATGACA
	ACAACTCGCA	ACCGAAATAT	GACCATTCTT	AAATATATTG	CGTATACTGT
1051	CTAAACAGGC	TTTTTCCAGT	AATTATGATT	CAGGTGTTTA	TTCATATTTA
	GATTTGTCCG	AAAAAGGTCA	TTAATACTAA	GTCCACAAAT	AAGTATAAAT
1101	ACCCCTTATT	TATCACACGG	TCGGTATTTT	AAACCATTAA	ATTTAGGTCA
	TGGGGAATAA	ATAGTGTGCC	AGCCATAAAG	TTTGGTAATT	TAAATCCAGT
1151	GAAGATGAAA	TTAACTAAAA	TATATTTGAA	AAAGTTTTCT	CGCGTTCTTT
	CTTCTACTTT	AATTGATTTT	ATATAAACTT	TTTCAAAAGA	GCGCAAGAAA
1201	GTCTTGCGAT	AGGATTTGCA	TCAGCATTTA	CATATAGTTA	TATAACCCAA
	CAGAACGCTA	TCCTAAACGT	AGTCGTAAAT	GTATATCAAT	ATATTGGGTT
1251	CCTAAGCCGG	AGGTAAAAAA	GGTAGTCTCT	CAGACCTATG	ATTTTGATAA
	GGATTGCGCC	TCCAATTTTT	CCATCAGAGA	GTCTGGATAC	TAAAACATT
1301	ATTCACTATT	GA CTCTTCTC	AGCGTCTTAA	TCTAAGCTAT	CGCTATGTTT
	TAAGTGATAA	CTGAGAAGAG	TCGCAGAATT	AGATTGATA	GCGATACAAA
1351	TCAAGGATTC	TAAGGGAAAA	TTAATTAATA	GCGACGATTT	ACAGAAAGCA
	AGTTCCTAAG	ATTCCCTTTT	AATTAATTAT	CGCTGCTAAA	TGTCTTCGTT
1401	GGTTATTCCA	TCACATATAT	TGATTTATGT	ACTGTTTCAA	TTAAAAAAGG
	CCAATAAGGT	AGTGTATATA	ACTAAATACA	TGACAAAGTT	AATTTTTTCC
1451	TAATTCAAAT	GAAATTGTTA	AATGTAATTA	ATTTTGTTTT	CTTGATGTTT
	ATTAAGTTTA	CTTTAACAAT	TTACATTAAT	TAAAACAAAA	GA ACTACAAA
1501	GTTTCATCAT	CTTCTTTTGC	TCAAGTAATT	GAAATGAATA	ATTCGCCTCT
	CAAAGTAGTA	GAAGAAAACG	AGTTCATTAA	CTTTACTTAT	TAAGCGGAGA
1551	GCGCGATTTT	GTGACTTGGT	ATTCAAAGCA	AACAGGTGAA	TCTGTTATTG
	CGCGCTAAAG	CACTGAACCA	TAAGTTTCGT	TTGTCCACTT	AGACAATAAC
1601	TCTCACCTGA	TGTTAAAGGT	ACAGTGA CTG	TATATTCCTC	TGACGTTAAG
	AGAGTG GACT	ACAATTTCCA	TGTC ACTGAC	ATATAAGGAG	ACTGCAATTC

27/39

1651 CCTGAAAATT TACGCAATTT CTTTATCTCT GTTTTACGTG CTAATAATTT
 GGACTTTTAA ATGCGTTAAA GAAATAGAGA CAAAATGCAC GATTATTAAA
 1701 TGATATGGTT GGCTCTAATC CTTCCATAAT TCAGAAATAT AACCCAAATA
 ACTATACCAA CCGAGATTAG GAAGGTATTA AGTCTTTATA TTGGGTTTAT
 1751 GTCAGGATTA TATTGATGAA TTGCCATCAT CTGATATTCA GGAATATGAT
 CAGTCCTAAT ATAACTACTT AACGGTAGTA GACTATAAGT CCTTATACTA
 1801 GATAATTCCG CTCCTTCTGG TGGTTTCTTT GTTCCGCAAA ATGATAATGT
 CTATTAAGGC GAGGAAGACC ACCAAAGAAA CAAGGCGTTT TACTATTACA
 1851 TACTCAAACA TTTAAAATTA ATAACGTTTCG CGCAAAGGAT TTAATAAGGG
 ATGAGTTTGT AAATTTTAAT TATTGCAAGC GCGTTTCCTA AATTATTCCC
 1901 TTGTAGAATT GTTTGTTAAA TCTAATACAT CTAAATCCTC AAATGTATTA
 AACATCTTAA CAAACAATTT AGATTATGTA GATTTAGGAG TTTACATAAT
 1951 TCTGTTGATG GTTCTAACTT ATTAGTAGTT AGCGCCCCTA AAGATATTTT
 AGACAACACT CAAGATTGAA TAATCATCAA TCGCGGGGAT TTCTATAAAA
 2001 AGATAACCTT CCGCAATTTT TTTCTACTGT TGATTTGCCA ACTGACCAGA
 TCTATTGGAA GGCGTTAAAG AAAGATGACA ACTAAACGGT TGACTGGTCT
 2051 TATTGATTGA AGGATTAATT TTCGAGGTTC AGCAAGGTGA TGCTTTAGAT
 ATAACCTAAT TCCTAATTAA AAGCTCCAAG TCGTTCCTACT ACGAAATCTA
 2101 TTTTCCTTTG CTGCTGGCTC TCAGCGCGGC ACTGTTGCTG GTGGTGTTAA
 AAAAGGAAAC GACGACCGAG AGTCGCGCCG TGACAACGAC CACCACAATT
 2151 TACTGACCGT CTAACCTCTG TTTTATCTTC TGCGGGTGGT TCGTTCGGTA
 ATGACTGGCA GATTGGAGAC AAAATAGAAG ACGCCCACCA AGCAAGCCAT
 2201 TTTTAAACGG CGATGTTTTA GGGCTATCAG TTCGCGCATT AAAGACTAAT
 AAAAATGGCC GCTACAAAAT CCCGATAGTC AAGCGCGTAA TTTCTGATTA
 2251 AGCCATTCAA AAATATTGTC TGTGCCTCGT ATTCTTACGC TTTCAGGTCA
 TCGGTAAGTT TTTATAACAG ACACGGAGCA TAAGAATGCG AAAGTCCAGT
 2301 GAAGGGTTCT ATTTCTGTTG GCCAGAATGT CCCTTTTATT ACTGGTCGTG
 CTTCCCAAGA TAAAGACAAC CGGTCTTACA GGGAAAATAA TGACCAGCAC
 2351 TAACTGGTGA ATCTGCCAAT GTAAATAATC CATTTACAGC AATTGAGCGT
 ATTGACCACT TAGACGGTTA CATTTATTAG GTAAAGTCTG TTAACGCA
 2401 CAAAATGTTG GTATTTCTAT GAGTGTTTTT CCCGTTGCAA TGGCTGGCGG
 GTTTTACAAC CATAAAGATA CTCACAAAAA GGGCAACGTT ACCGACCGCC
 2451 TAATATTGTT TTAGATATAA CCAGTAAGGC CGATAGTTTG AGTTCTTCTA
 ATTATAACAA AATCTATATT GGTCATTCCG GCTATCAAAC TCAAGAAGAT

28/39

2501 CTCAGGCAAG TGATGTTATT ACTAATCAAA GAAGTATTGC GACAACGGTT
GAGTCCGTTT ACTACAATAA TGATTAGTTT CTTCATAACG CTGTTGCCAA

2551 AATTTGCGTG ATGGTCAGAC TCTTTTGCTC GGTGGCCTCA CTGATTACAA
TTAAACGCAC TACCAGTCTG AGAAAACGAG CCACCGGAGT GACTAATGTT

2601 AAACACTTCT CAAGATTCTG GTGTGCCGTT CCTGTCTAAA ATCCCTTTAA
TTTGTGAAGA GTTCTAAGAC CACACGGCAA GGACAGATTT TAGGGAAATT

2651 TCGGCCTCCT GTTTAGCTCC CGTTCTGATT CTAACGAGGA AAGCACGTTG
AGCCGGAGGA CAAATCGAGG GCAAGACTAA GATTGCTCCT TTCGTGCAAC

2701 TACGTGCTCG TCAAAGCAAC CATAGTACGC GCCCTGTAGC GGCGCATTAA
ATGCACGAGC AGTTTCGTTG GTATCATGCG CGGGACATCG CCGCGTAATT

2751 GCGCGGCGGG TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC
CGCGCCGCCC ACACCACCAA TGCGCGTCGC ACTGGCGATG TGAACGGTCG

2801 GCCCTAGCGC CCGCTCCTTT CGCTTCTTC CCTTCCTTTC TCGCCACGTT
CGGGATCGCG GGCGAGGAAA GCGAAAGAAG GGAAGGAAAG AGCGGTGCAA

BamHI

2851 CTCCGGCTTT CCCCGTCAAG CTCTAAATCG GGGGATCCCT TTAGGGTTCC
GAGGCCGAAA GGGGCAGTTC GAGATTTAGC CCCCTAGGGA AATCCCAAGG

2901 GATTTAGTGC TTTACGGCAC CTCGACCTCC AAAAATTGA TTTGGGTGAT
CTAAATCACG AAATGCCGTG GAGCTGGAGG TTTTGAAGT AAACCCACTA

2951 GGTTACAGTA GTGGGCCATC GCCCTAATAG ACGGTTTTTC GCCCTTTGAC
CCAAGTGCAT CACCCGGTAG CGGGATTATC TGCCAAAAAG CGGGAAACTG

3001 GTTGGAGTCC ACGTTCTTTA ATAGTGGACT CTTGTTCCAA ACTGGAACAA
CAACCTCAGG TGCAAGAAAT TATCACCTGA GAACAAGGTT TGACCTTGTT

3051 CACTCAACCC TATCTCGGTC TATTCTTTTG ATTTATAAGG GATTTTGCCG
GTGAGTTGGG ATAGAGCCAG ATAAGAAAAC TAAATATTCC CTAAAACGGC

3101 ATTTCCGCCT ATTGGTTAAA AAATGAGCTG ATTTAACAAA AATTTAACGC
TAAAGCCGGA TAACCAATTT TTTACTCGAC TAAATTGTTT TTAAATTGCG

3151 GAATTTTAAC AAAATATTAA CGTTTACAAT TTAAATATTT GCTTATACAA
CTTAAAATTG TTTTATAATT GCAAATGTTA AATTTATAAA CGAATATGTT

3201 TCTTCCTGTT TTTGGGGCTT TTCTGATTAT CAACCGGGGT ACATATGATT
AGAAGGACAA AAACCCCGAA AAGACTAATA GTTGCCCCA TGTATACTAA

ClaI

3251 GACATGCTAG TTTTACGATT ACCGTTTCATC GATTCTCTTG TTTGCTCCAG
CTGTACGATC AAAATGCTAA TGGCAAGTAG CTAAGAGAAC AAACGAGGTC

29/39

3301 ACTCTCAGGC AATGACCTGA TAGCCTTTTT AGACCTCTCA AAAATAGCTA
 TGAGAGTCCG TTAGTGACT ATCGGAAAAA TCTGGAGAGT TTTTATCGAT
 3351 CCCTCTCCGG CATGAATTTA TCAGCTAGAA CGGTTGAATA TCATATTGAT
 GGGAGAGGCC GTACTTAAAT AGTCGATCTT GCCAACTTAT AGTATAACTA
 3401 GGTGATTTGA CTGTCTCCGG CCTTTCTCAC CCGTTTGAAT CTTTACCTAC
 CCACTAACT GACAGAGGCC GGAAAGAGTG GGCAAACCTTA GAAATGGATG
 3451 ACATTACTCA GGCATTGCAT TTAAAATATA TGAGGGTTCT AAAAATTTTT
 TGTAATGAGT CCGTAACGTA AATTTTATAT ACTCCCAAGA TTTTAAAAA
 3501 ATCCTTGCGT TGAAATAAAG GCTTCTCCCG CAAAAGTATT ACAGGGTCAT
 TAGGAACGCA ACTTTATTTT CGAAGAGGGC GTTTTCATAA TGTCCCAGTA
 3551 AATGTTTTTG GTACAACCGA TTTAGCTTTA TGCTCTGAGG CTTTATTGCT
 TTACAAAAAC CATGTTGGCT AAATCGAAAT ACGAGACTCC GAAATAACGA
 3601 TAATTTTGCT AATTCTTTGC CTTGCCTGTA TGATTTATTG GATGTTAACG
 ATTAAACGA TTAAGAAACG GAACGGACAT ACTAAATAAC CTACAATTGC
 3651 CTACTACTAT TAGTAGAATT GATGCCACCT TTTCAGCTCG CGCCCCAAAT
 GATGATGATA ATCATCTTAA CTACGGTGGA AAAGTCGAGC GCGGGGTTTA
 3701 GAAAATATAG CTAAACAGGT TATTGACCAT TTGCGAAATG TATCTAATGG
 CTTTTATATC GATTTGTCCA ATAAGTGGTA AACGCTTTAC ATAGATTACC
 3751 TCAAACATAA TCTACTCGTT CGCAGAATTG GGAATCAACT GTTACATGGA
 AGTTTGATTT AGATGAGCAA GCGTCTTAAC CCTTAGTTGA CAATGTACCT
 3801 ATGAAACTTC CAGACACCGT ACTTTAGTTG CATATTTAAA ACATGTTGAG
 TACTTTGAAG GTCTGTGGCA TGAAATCAAC GTATAAATTT TGTACAACCTC
 3851 CTACAGCACC AGATCCAGCA ATTAAGCTCT AAGCCATCCG CAAAAATGAC
 GATGTCGTGG TCTAGGTCGT TAATTCGAGA TTCGGTAGGC GTTTTTACTG
 3901 CTCTTATCAA AAGGAGCAAT TAAAGGTACT CTCTAATCCT GACCTGTTGG
 GAGAATAGTT TTCCTCGTTA ATTTCCATGA GAGATTAGGA CTGGACAACC
 3951 AGTTTGCTTC CGGTCTGGTT CGCTTTGAAG CTCGAATTAA AACGCGATAT
 TCAAACGAAG GCCAGACCAA GCGAACTTC GAGCTTAATT TTGCGCTATA
 4001 TTGAAGTCTT TCGGGCTTCC TCTTAATCTT TTTGATGCAA TCCGCTTTGC
 AACTTCAGAA AGCCCGAAGG AGAATTAGAA AACTACGTT AGGCGAAACG
 4051 TTCTGACTAT AATAGTCAGG GTAAAGACCT GATTTTTGAT TTATGGTCAT
 AAGACTGATA TTATCAGTCC CATTTCTGGA CTAAAACTA AATACCAGTA
 4101 TCTCGTTTTT TGAAGTGTG AAAGCATTTG AGGGGGATTC AATGAATATT
 AGAGCAAAAG ACTTGACAAA TTTCGTAAAC TCCCCCTAAG TTACTTATAA

30/39

4151 TATGACGATT CCGCAGTATT GGACGCTATC CAGTCTAAAC ATTTTACTAT
 ATACTGCTAA GGCGTCATAA CCTGCGATAG GTCAGATTG TAAAATGATA
 4201 TACCCCTCTT GGCAAACTT CTTTTGCAA AGCCTCTCGC TATTTTGTG
 ATGGGGGAGA CCGTTTGTAA GAAAACGTT TCGGAGAGCG ATAAAAACAA
 4251 TTTATCGTCG TCTGGTAAAC GAGGGTTATG ATAGTGTGTC TCTTACTATG
 AAATAGCAGC AGACCATTG CTCCCAATAC TATCACAACG AGAATGATAC
 4301 CCTCGTAATT CCTTTTGGCG TTATGTATCT GCATTAGTTG AATGTGGTAT
 GGAGCATTAA GGAAAACCGC AATACATAGA CGTAATCAAC TTACACCATA
 4351 TCCTAAATCT CAACTGATGA ATCTTCTAC CTGTAATAAT GTTGTTCGGT
 AGGATTTAGA GTTGACTACT TAGAAAGATG GACATTATTA CAACAAGGCA
 4401 TAGTTCGTTT TATTAACGTA GATTTTTCTT CCCAACGTCC TGAAGGTAT
 ATCAAGCAAA ATAATTGCAT CTAAAAAGAA GGGTTGCAGG ACTGACCATA
 4451 AATGAGCCAG TTCTTAAAT CGCATAAGGT AATTCACAAT GATTAAAGTT
 TTAAGCGTC AAGAATTTTA GCGTATTCCA TTAAGTGTTA CTAATTTCAA
 4501 GAAATTAAAC CATCTCAAGC GCAATTCAT ACCCGTTCTG GTGTTTCTCG
 CTTTAATTTG GTAGAGTTTCG CGTTAAGTGA TGGGCAAGAC CACAAAGAGC
 4551 TCAGGGCAAG CCTTATTCAC TGAATGAGCA GCTTTGTTAC GTTGATTG
 AGTCCCGTTC GGAATAAGTG ACTTACTCGT CGAAACAATG CAACTAAACC
 4601 GTAATGAATA TCCGGTGCTT GTCAAGATTA CTCTTGATGA AGGTCAGCCA
 CATTACTTAT AGGCCACGAA CAGTTCTAAT GAGAACTACT TCCAGTCGGT
 4651 GCCTATGCGC CTGGTCTGTA CACCGTGCAT CTGTCCTCGT TCAAAGTTGG
 CGGATACGCG GACCAGACAT GTGGCACGTA GACAGGAGCA AGTTTCAACC
 4701 TCAGTTCGGT TCTCTTATGA TTGACCGTCT GCGCCTCGTT CCGGCTAAGT
 AGTCAAGCCA AGAGAATACT AACTGGCAGA CGCGGAGCAA GGCCGATTCA
 4751 AACATGGAGC AGGTCGCGGA TTTCGACACA ATTTATCAGG CGATGATACA
 TTGTACCTCG TCCAGCGCCT AAAGCTGTGT TAAATAGTCC GCTACTATGT
 4801 AATCTCCGTT GTACTTTGTT TCGCGCTTGG TATAATCGCT GGGGGTCAAA
 TTAGAGGCAA CATGAAACAA AGCGCAACC ATATTAGCGA CCCCAGTTT
 4851 GATGAGTGTT TTAGTGTATT CTTTCGCCTC TTTCGTTTTA GGTGGTGCC
 CTACTCACAA AATCACATAA GAAAGCGGAG AAAGCAAAAT CCAACCACGG
 4901 TTCGTAGTGG CATTACGTAT TTTACCGTT TAATGGAAAC TTCCTCATGC
 AAGCATCACC GTAATGCATA AAATGGGCAA ATTACCTTTG AAGGAGTACG
 4951 GTAAGTCTTT AGTCCTCAAA GCCTCCGTAG CCGTTGCTAC CCTCGTTCCG
 CATTACAGAA TCAGGAGTTT CGGAGGCATC GGCAACGATG GGAGCAAGGC

31/39

5001 ATGCTGTCTT TCGCTGCTGA GGGTGACGAT CCCGCAAAG CGGCCTTTGA
 TACGACAGAA AGCGACGACT CCCACTGCTA GGGCGTTTTT GCCCGAAACT
 5051 CTCCCTGCAA GCCTCAGCGA CCGAATATAT CGGTTATGCG TGGGCGATGG
 GAGGGACGTT CGGAGTCGCT GGCTTATATA GCCAATACGC ACCCGCTACC
 5101 TTGTTGTCAT TGTCGGCGCA ACTATCGGTA TCAAGCTGTT TAAGAAATTC
 AACAACAGTA ACAGCCGCGT TGATAGCCAT AGTTCGACAA ATTCTTTAAG
 5151 ACCTCGAAAG CAAGCTGATA AAGGAGGTTT CTCGATCGAG ACGTTGGGTG
 TGGAGCTTTC GTTCGACTAT TTCCTCCAAA GAGCTAGCTC TGCAACCCAC
 5201 AGGTTCCAAC TTTCACCATA ATGAAATAAG ATCACTACCG GCGGTATTTT
 TCCAAGGTTG AAAGTGGTAT TACTTTATTTC TAGTGATGGC CCGCATAAAA
 5251 TTGAGTTATC GAGATTTTCA GGAGCTAAGG AAGCTAAAAT GGAGAAAAAA
 AACTCAATAG CTCTAAAAGT CCTCGATTCC TTCGATTTTA CCTCTTTTTT
 5301 ATCACTGGAT ATACCACCGT TGATATATCC CAATGGCATC GTAAAGAACA
 TAGTGACCTA TATGGTGGCA ACTATATAGG GTTACCGTAG CATTTCTTGT
 5351 TTTTGAGGCA TTTCAGTCAG TTGCTCAATG TACCTATAAC CAGACCGTTC
 AAAACTCCGT AAAGTCAGTC AACGAGTTAC ATGGATATTG GTCTGGCAAG
 5401 AGCTGGATAT TACGGCCTTT TTAAAGACCG TAAAGAAAAA TAAGCACAAG
 TCGACCTATA ATGCCGGAAG AATTTCTGGC ATTTCTTTTT ATTCTGTGTC
 5451 TTTTATCCGG CCTTTATTCA CATTCTTGCC CGCCTGATGA ATGCTCATCC
 AAAATAGGCC GGAAATAAGT GTAAGAACGG GCGGACTACT TACGAGTAGG
 5501 GGAGTTCCGT ATGGCAATGA AAGACGGTGA GCTGGTGATA TGGGATAGTG
 CCTCAAGGCA TACCGTTACT TTCTGCCACT CGACCACTAT ACCCTATCAC
 5551 TTCACCCTTG TTACACCGTT TTCCATGAGC AAAGTGAAAC GTTTTCATCG
 AAGTGGGAAC AATGTGGCAA AAGGTACTCG TTTGACTTTG CAAAAGTAGC
 5601 CTCTGGAGTG AATACCACGA CGATTTCCGG CAGTTTCTAC ACATATATTC
 GAGACCTCAC TTATGGTGCT GCTAAAGGCC GTCAAAGATG TGTATATAAG
 5651 GCAAGATGTG GCGTGTTACG GTGAAAACCT GGCCTATTTT CCTAAAGGGT
 CGTTCTACAC CGCACAATGC CACTTTTGGA CCGGATAAAG GGATTTCCCA
 5701 TTATTGAGAA TATGTTTTTC GTCTCAGCCA ATCCCTGGGT GAGTTTCACC
 AATAACTCTT ATACAAAAG CAGAGTCGGT TAGGGACCCA CTCAAAGTGG
 5751 AGTTTTGATT TAAACGTAGC CAATATGGAC AACTTCTTCG CCCCCGTTTT
 TCAAACTAA ATTTGCATCG GTTATACCTG TTGAAGAAGC GGGGGCAAAA
 5801 CACTATGGGC AAATATTATA CGCAAGGCGA CAAGGTGCTG ATGCCGCTGG
 GTGATACCCG TTTATAATAT GCGTTCGCT GTTCCACGAC TACGGCGACC

32/39

5851	CGATTCAGGT	TCATCATGCC	GTTTGTGATG	GCTTCCATGT	CGGCAGAATG
	GCTAAGTCCA	AGTAGTACGG	CAAACACTAC	CGAAGGTACA	GCCGTCTTAC
5901	CTTAATGAAT	TACAACAGTA	CTGCGATGAG	TGGCAGGGCG	GGGCGTAATT
	GAATTACTTA	ATGTTGTCAT	GACGCTACTC	ACCGTCCCCG	CCCGCATTAA
5951	TTTTTAAGGC	AGTTATTGGT	GCCCTTAAAC	GCCTGGTGCT	AGCCTGAGGC
	AAAAATTCCG	TCAATAACCA	CGGGAATTTG	CGGACCACGA	TCGGACTCCG
6001	CAGTTTGCTC	AGGCTCTCCC	CGTGGAGGTA	ATAATTGCTC	GACCGATAAA
	GTCAAACGAG	TCCGAGAGGG	GCACCTCCAT	TATTAACGAG	CTGGCTATTT
6051	AGCGGCTTCC	TGACAGGAGG	CCGTTTTGTT	TTGCAGCCCA	CCTCAACGCA
	TCGCCGAAGG	ACTGTCCTCC	GGCAAACAA	AACGTCGGGT	GGAGTTGCGT
6101	ATTAATGTGA	GTTAGCTCAC	TCATTAGGCA	CCCCAGGCTT	TACACTTTAT
	TAATTACACT	CAATCGAGTG	AGTAATCCGT	GGGGTCCGAA	ATGTGAAATA
6151	GCTTCCGGCT	CGTATGTTGT	GTGGAATTGT	GAGCGGATAA	CAATTTCACA
	CGAAGGCCGA	GCATACAACA	CACCTTAACA	CTCGCCTATT	GTTAAAGTGT
6201	CAGGAAACAG	CTATGACCAT	GATTACGAAT	TTCTAGATAA	CGAGGGCAAA
	GTCCTTTGTC	GATACTGGTA	CTAATGCTTA	AAGATCTATT	GCTCCCGTTT
6251	AAATGAAAAA	GACAGCTATC	GCGATTGCAG	TGGCACTGGC	TGGTTTTGCT
	TTTACTTTTT	CTGTCGATAG	CGCTAACGTC	ACCGTGACCG	ACCAAAGCGA
6301	ACCGTAGCGC	AGGCCGACTA	CAAAGATGTC	GACTGTATTG	TTTATCATGC
	TGGCATCGCG	TCCGGCTGAT	GTTTCTACAG	CTGACATAAC	AAATAGTACG
			BamHI	EcoRI	
			~~~~~		
6351	TCATTATCTT	GTTGCTAAGT	GTGGTGGTGG	AGGATCCGAA	TTCAATGCTG
	AGTAATAGAA	CAACGATTCA	CACCACCACC	TCCTAGGCTT	AAGTTACGAC
6401	GCGGCGGCTC	TGGTGGTGGT	TCTGGTGGCG	GCTCTGAGGG	TGGTGGCTCT
	CGCCGCCGAG	ACCACCACCA	AGACCACCGC	CGAGACTCCC	ACCACCAGAA
6451	GAGGGTGGCG	GTTCTGAGGG	TGGCGGCTCT	GAGGGAGGCG	GTTCCGGTGG
	CTCCCACCGC	CAAGACTCCC	ACCGCCGAGA	CTCCCTCCGC	CAAGGCCACC
6501	TGGCTCTGGT	TCCGGTGATT	TTGATTATGA	AAAGATGGCA	AACGCTAATA
	ACCGAGACCA	AGGCCACTAA	AACTAATACT	TTTCTACCGT	TTGCGATTAT
6551	AGGGGGCTAT	GACCGAAAAAT	GCCGATGAAA	ACGCGCTACA	GTCTGACGCT
	TCCCCCGATA	CTGGCTTTTA	CGGCTACTTT	TGCGCGATGT	CAGACTGCGA

33/39

Clai

-----

6601 AAAGGCAAAC TTGATTCTGT CGCTACTGAT TACGGTGCTG CTATCGATGG  
TTTCCGTTTG AACTAAGACA GCGATGACTA ATGCCACGAC GATAGCTACC

6651 TTTCATTGGT GACGTTTCCG GCCTTGCTAA TGGTAATGGT GCTACTGGTG  
AAAGTAACCA CTGCAAAGGC CGGAACGATT ACCATTACCA CGATGACCAC

6701 ATTTTGCTGG CTCTAATTCC CAAATGGCTC AAGTCGGTGA CGGTGATAAT  
TAAAACGACC GAGATTAAGG GTTTACCGAG TTCAGCCACT GCCACTATTA

6751 TCACCTTTAA TGAATAATTT CCGTCAATAT TTACCTTCCC TCCCTCAATC  
AGTGGAATTT ACTTATTAAA GGCAGTTATA AATGGAAGGG AGGGAGTTAG

6801 GGTGAATGT CGCCCTTTTG TCTTTGGCGC TGGTAAACCA TATGAATTTT  
CCAACCTTACA GCGGGAAAAC AGAAACCGCG ACCATTGGT ATACTTAAAA

6851 CTATTGATTG TGACAAAATA AACTTATTCC GTGGTGTCTT TGCGTTTCTT  
GATAACTAAC ACTGTTTAT TTGAATAAGG CACCACAGAA ACGCAAAGAA

6901 TTATATGTTG CCACCTTTAT GTATGTATTT TCTACGTTTG CTAACATACT  
AATATACAAC GGTGGAAATA CATACATAAA AGATGCAAAC GATTGTATGA

HindIII

6951 GCGTAATAAG GAGTCTTGAT A  
CGCATTATTC CTCAGAACTA T

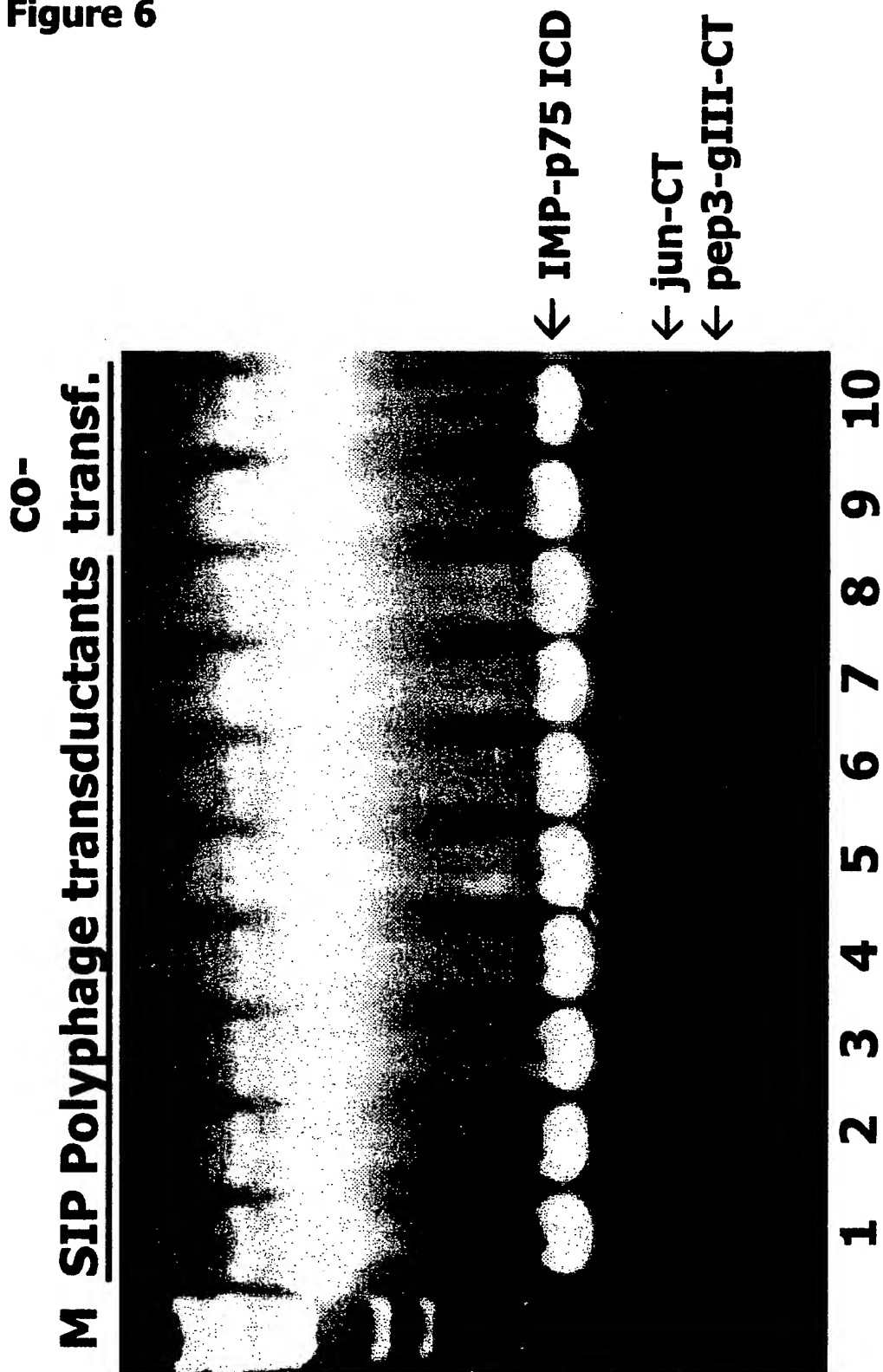
34/39

**Figure 5**



35/39

**Figure 6**



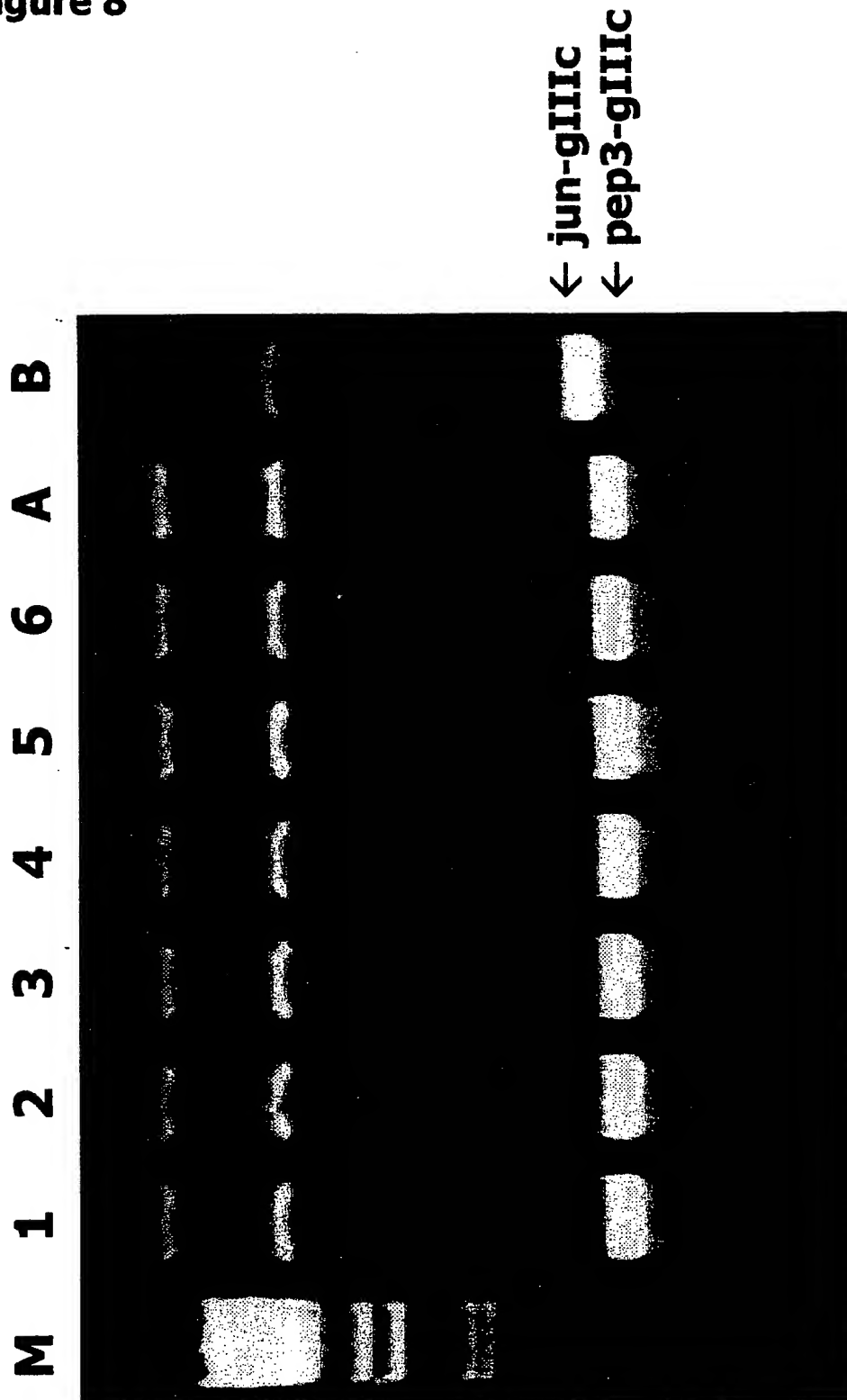
36/39

**Figure 7**

<b>pep3/p75ICD</b>		<b>dilution factor</b>		<b>transductants</b>	
		<b>jun/p75ICD</b>		<b>(t.u./ml)*</b>	
<b>1</b>	<b>pos. control</b>	<b>-</b>		<b>6 x 10⁵</b>	
<b>-</b>	<b>neg. control</b>	<b>1</b>		<b>0</b>	
<b>1</b>		<b>10²</b>		<b>1.2 x 10⁴</b>	
<b>1</b>		<b>10³</b>		<b>8.6 x 10²</b>	
<b>1</b>		<b>10⁴</b>		<b>1.2 x 10²</b>	
<b>1</b>		<b>10⁵</b>		<b>12[#]</b>	
<b>1</b>		<b>10⁶</b>		<b>1.2[#]</b>	
<b>1</b>		<b>10⁷</b>		<b>0.12[#]</b>	

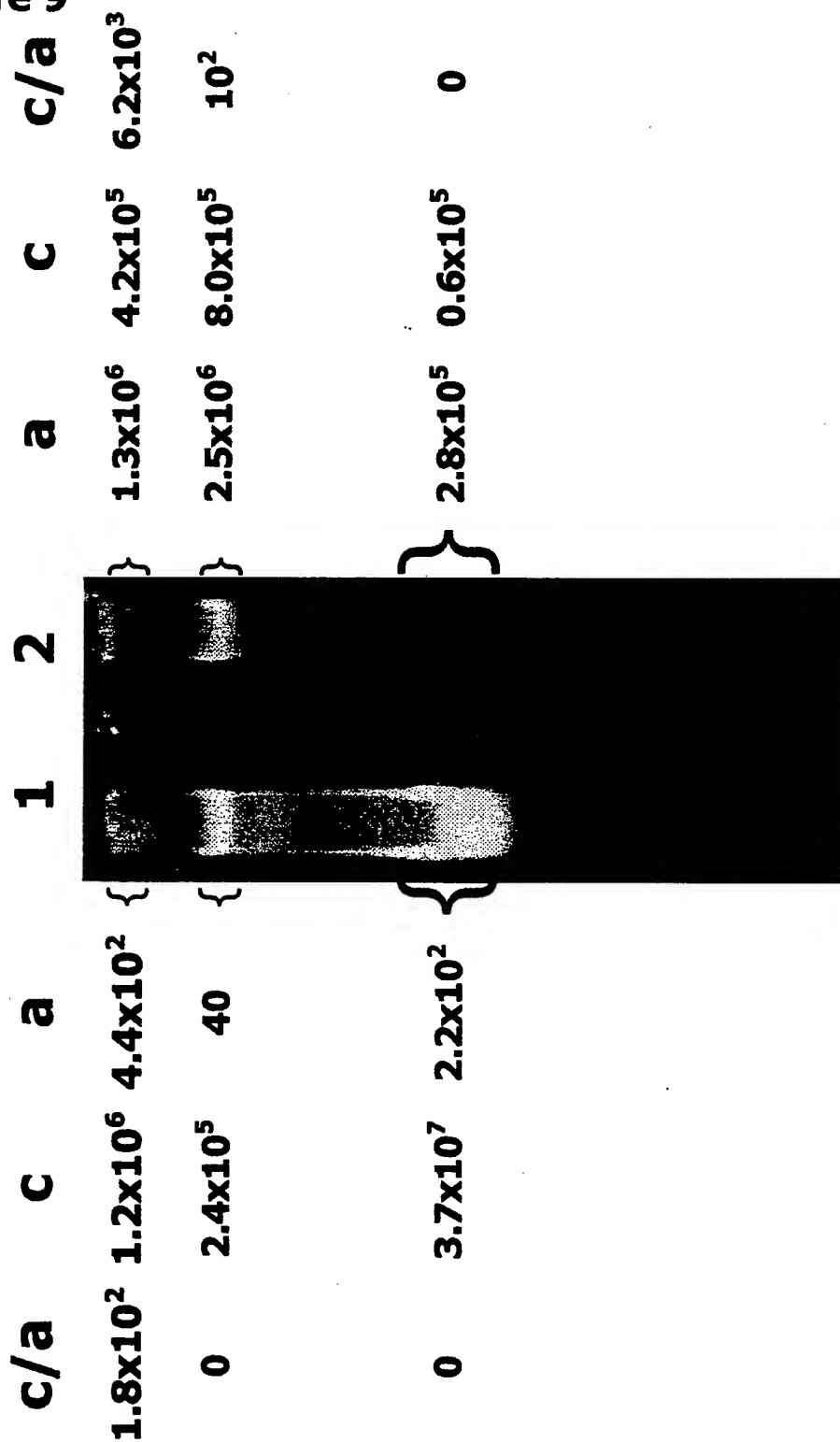
37/39

**Figure 8**



38/39

Figure 9







**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**